

Rekonstruktion von abgeleiteten Variablen im Mikrozensus 2005 im Vergleich zu den Mikrozensen bis 2004

John, Kristina

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

John, K. (2008). *Rekonstruktion von abgeleiteten Variablen im Mikrozensus 2005 im Vergleich zu den Mikrozensen bis 2004*. (GESIS-Methodenberichte, 7/2008). Mannheim: GESIS-ZUMA. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-206902>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

GESIS-Methodenbericht 7/2008
**Rekonstruktion von abgeleiteten Variablen
im Mikrozensus 2005 im Vergleich zu
den Mikrozensen bis 2004**

Kristina John

Juni 2008

Kristina John
GESIS
German Microdata Lab (GML)
Quadrat B2,1
Postfach 12 21 55
68072 Mannheim
Telefon: (0621) 1246-252
Telefax: (0621) 1246-100
E-Mail: Kristina.John@gesis.org

GESIS Methodenbericht 7/2008

GESIS-Methodenberichte

Die GESIS ist ein Institut der Leibniz-Gemeinschaft.

ISSN:

Herausgeber, Druck und Vertrieb:

1865-7575

GESIS

Postfach 12 21 55, 68072 Mannheim

Printed in Germany

Inhaltsverzeichnis

1	Einleitung.....	1
2	Neuerungen im Mikrozensus 2005	2
2.1	Konzepte und Definitionen.....	3
2.1.1	<i>Haushalt</i>	5
2.1.2	<i>Bezugsperson des Haushalts</i>	6
2.1.3	<i>Lebensform</i>	6
2.1.4	<i>Familie</i>	7
2.1.5	<i>Abgrenzungen der Bevölkerungsbegriffe</i>	8
3	Rekonstruktion von abgeleiteten Variablen.....	9
3.1	Replikation der Variable <i>EF664</i> "Zahl der Erwerbstätigen im Haushalt"	11
3.2	Replikation der Variable <i>EF713</i> "Zahl der nichtehelichen Lebensgemeinschaften im Haushalt"	13
3.3	Replikation der Variable <i>EF736</i> "Erwerbstyp: Haupteinkommensbezieher"	14
3.4	Replikation der Variable <i>EF758</i> "Erwerbstyp: Haushaltsbezugsperson"	15
3.5	Replikation der Variable <i>EF770</i> "Zahl der ledigen Kinder in der Familie/Lebensform insgesamt (ohne Altersbegrenzung)"	15
3.6	Replikation der Variable <i>EF824</i> "Erwerbstyp der Bezugsperson in der Familie/Lebensform"	16
3.7	Replikation der Variable <i>EF848</i> "Erwerbstyp des Ehe- oder Lebenspartners"	16
3.8	Replikation der Variable <i>EF866</i> "Zahl der ledigen Kinder insgesamt in der Familie"	17
4	Diskrepanzen im Mikrozensus 2005	18
	Literatur.....	20
	Anhang:	
	Erläuterungen zu Tabelle 2	21
	STATA do-Files	23
	SPSS sps-Files	39

Zusammenfassung

Neben Informationen auf der Mikro- oder Makroebene sind für eigene wissenschaftliche Fragestellungen oftmals sozio-ökonomische Kontextinformationen zu den gesellschaftlichen Mesoebenen Haushalt, Familie oder Lebensform von Relevanz. Diese Informationen liegen im Mikrozensus als sogenannte abgeleitete Variablen vor. Obgleich der Mikrozensus 2005 mit 146 abgeleiteten Variablen ein enormes Analysepotenzial birgt, sind abhängig von der eigenen Forschungsfrage jedoch teilweise andere Rekodierungen oder Typisierungen nötig.

Der vorliegende Bericht hat drei Zielsetzungen:

1. Zum Einen soll er die für die Rekonstruktion mithilfe der Daten des Mikrozensus 2005 relevanten Konzepte und Definitionen „Bezugsperson des Haushalts“, „Haushalt“, „Lebensform“, „Familie“ und die Abgrenzungen der Bevölkerungsbegriffe darstellen.
2. Zum Anderen sollen diese Konzepte mit denen der vorherigen Mikrozensen bis 2004 verglichen und allgemeine Neuerungen des Mikrozensus 2005 benannt werden.
3. Zudem soll der Bericht anhand von acht Variablen beispielhaft zeigen, wie derartige abgeleitete Variablen mithilfe der Statistik-Programme STATA und SPSS selbst konstruiert werden können. Dabei soll illustriert werden, welche Konzepte und Identifikatoren für die jeweilige Analyseeinheit, d.h. den Haushalt, die Familie oder die Lebensform, benötigt werden.

1 Einleitung

Der Mikrozensus erhebt 1% der Personen und Haushalte in Deutschland.¹ Aufgrund der Auskunftspflicht auf Basis des Mikrozensus-Gesetzes und der resultierenden hohen Ausschöpfungsquote aus einer Grundgesamtheit von 82.5 Millionen Personen im Jahr 2005 birgt der Mikrozensus eine solide Datengrundlage. Neben Informationen zu Bevölkerung und Arbeitsmarkt sind differenzierte Analysen zu Haushalten, Familien oder Lebensformen als Untereinheiten des Haushaltes und zu Personen möglich.

Der familiensoziologische Kontext der drei Analyseebenen Haushalt, Familie und Lebensform wird durch so genannte *abgeleitete Variablen (derived variables)*, einst auch Bandsatzerweiterungen genannt, vom Statistischen Bundesamt post-hoc mithilfe der im Mikrozensus direkt erhobenen Merkmale generiert.²

Im vorliegenden Bericht wird anhand von acht ausgewählten Beispielen dargestellt, wie derartige abgeleitete Variablen selbst erzeugt werden können. Die ausführliche Syntax der Statistikprogramme STATA (Version 10) und SPSS (Version 16) ist in den Anhängen enthalten. Fokussiert man exemplarisch auf die Meso-Ebene Haushalt, sind somit Informationen zu beispielsweise der Anzahl der Kinder im Haushalt (*EF669*) oder zur Stellung im Beruf des Haupteinkommensbeziehers im Haushalt (*EF740*) vorhanden.

Ausgestattet mit diesem Wissen über die Rekonstruktion bietet sich der Forscherin oder dem Forscher in der Folge die Perspektive, eigene abgeleitete Variablen mit für die eigenen Analysen spezifischen Abgrenzungen zu generieren. Beispielsweise können komplett andere oder im Differenzierungsgrad variierende Abgrenzungen gewählt werden als die in der amtlichen Statistik verwendeten. Im Folgenden soll der Mikrozensus 2005 mit dem Mikrozensus 2004 verglichen werden, wobei insbesondere auf die Innovationen des Mikrozensus 2005 eingegangen wird. Bemerkenswert sind vor allem einige neue Variablen, die neuen Abgrenzungen bei Familien beziehungsweise Lebensformen und dass es nach der Umstellung auf das neue „Konzept der Lebensformen“ nur noch wenige abge-

1 Damit ist er die größte jährliche Haushaltsbefragung in Europa. Seine Repräsentativität ist durch sein Design als rotierende Panelstichprobe und zudem Flächen- bzw. Klumpenstichprobe, in der die Haushalte eines Auswahlbezirks vier Jahre lang befragt werden, gegeben. Jedes Jahr wird ein Viertel der Auswahlbezirke (Klumpen) ausgetauscht.

2 Kontextinformationen werden nur für die in einem Haushalt zusammen lebenden Personen, sei es in einer Familie oder Lebensgemeinschaft, gewonnen. Soziale Beziehungen zwischen Personen, wie Familienmitgliedern, die nicht zusammen in einem Haushalt wohnen, können nicht illustriert werden (Koresidenzprinzip). Eine Ausnahme bilden Fragen zu Migration. Nehmen wir an, eine Person mit Migrationshintergrund bejaht beispielsweise zunächst die freiwillig zu beantwortende Frage, ob die Eltern noch im Herkunftsland leben, (vgl. „Fragen zu Angehörigen im Herkunftsland“ Frage 102a im Fragebogen, S. 30) und verneint anschließend die Frage, ob der Vater, bzw. die Mutter auch in diesem Haushalt leben, (vgl. „Fragen zur Staatsangehörigkeit der Eltern“ Frage 103 bzw. 110 im Fragebogen S. 30ff.). Dann werden die Informationen über die sozialen Beziehungen der Familie oder Lebensgemeinschaft über die Grenzen des Haushaltes transferiert.

leitete Variablen zum vorherigen „traditionellen Familienkonzept“ gibt. Jedoch sind beide Konzepte mit dem Ziel der Schaffung der intertemporellen Vergleichbarkeit vorhanden: Falls Vergleiche zu Mikrozensen vor 2004 angestellt werden sollen, ist es daher sinnvoll zu wissen, wie auch nach dem „traditionellen Familienkonzept“, d.h. über das Ehe- und Abstammungsprinzip³, abgeleitete Variablen gebildet werden können. Im Mikrozensus 2005 wird eine Lebensgemeinschaft nur auf dem Merkmal beruhend, ob eine Eltern-Kind-Gemeinschaft vorliegt, als Familie deklariert. Zudem stehen die Integration der „nicht-konventionellen Lebensformen“ wie „Singles“, oder nicht-eheliche und gleichgeschlechtliche Lebensformen in die laufende Berichterstattung im Zentrum der Umstellung auf das „Lebensformenkonzept“ (vgl. Nöthen, 2005). Das folgende Kapitel beinhaltet eine deskriptive Darstellung der genannten Neuerungen.⁴

2 Neuerungen im Mikrozensus 2005

Ziel des vorliegenden Berichts ist es vor Allem, die Neuerungen des Mikrozensus 2005 im Vergleich zum Mikrozensus 2004 bezüglich der genannten Kontextinformationen darzustellen⁵:

Augenscheinlich haben einige Variablen auch einen neuen Namen erhalten. Beispielsweise wurden die Namen der für viele Rekonstruktionen von abgeleiteten Variablen benötigten Ausgangsvariablen wie Alter oder Erwerbstyp umbenannt. In Tabelle 1 sind die Namensänderungen bezüglich der verwendeten Ausgangsvariablen, die für die in diesem Bericht beispielhaft dargestellten abgeleiteten Variablen benötigt werden, aufgeführt.

Im Vergleich zum Mikrozensus 2004 wurde der Mikrozensus 2005 durch 161 neue Variablen bereichert. Von diesen neuen Variablen sind 90 als abgeleitete Variablen zu klassifizieren; insgesamt sind 146 abgeleitete Variablen vorhanden. Das volle Analysepotenzial ergibt sich aus 435 Variablen.

³ Der Familienbegriff bis einschließlich des Mikrozensus 2004 bezog sich sowohl auf Ehe- als auch Eltern-Kind-Gemeinschaften.

⁴ Mein besonderer Dank gilt Bernhard Schimpl-Neimanns für seine Unterstützung.

⁵ Der Mikrozensus 2005 unterscheidet sich jedoch auch methodisch und inhaltlich vom Mikrozensus 2004: Durch die Einführung des neuen Mikrozensusgesetzes zum 1. Januar 2005 wurde das Erhebungskonzept vom Berichtswochenkonzept auf eine unterjährige Erhebung umgestellt, sodass sowohl Jahresdurchschnitte als auch Quartals- und Monatsdurchschnitte ermittelt werden können. Durch die Aufnahme des Themenkomplexes „Migration und Integration“ in die Erhebung, bietet der Mikrozensus 2005 auch inhaltlich Analysepotenzial aktueller gesellschaftspolitischer Entwicklungen. (Statistisches Bundesamt, 2006a). Für eine Einführung in den Mikrozensus 2005 Scientific Use File sei auf Lechert und Schimpl-Neimanns (2007) verwiesen.

Tabelle 1: Änderungen hinsichtlich der Variablen-Bezeichnungen (EF: Eingabefeld)

Variablen	Bezeichnung im	
	Mikrozensus 2005	Mikrozensus 2004
Alter	EF44	EF30
Geschlecht	EF46	EF32
Stellung im Beruf	EF117	EF127
Bevölkerung: Privathaushalte	EF31	EF506
Erwerbstyp	EF29	EF504

Jedoch sind im Mikrozensus 2005 im Vergleich zum Mikrozensus 2004 weniger Typen von abgeleiteten Variablen auffindbar. Im Mikrozensus 2004 sind 89 Typen vorhanden, während im Mikrozensus 2005 81 Typen ersichtlich sind, jedoch sind letztere deutlich differenzierter: Betrachtet man beispielsweise den Variablen-Typ der Anzahl der Kinder im Haushalt eines bestimmten Alters, weisen die Alterskategorien im Mikrozensus 2004 eine größere Weite auf, da sie in 10 Kategorien unterteilt sind (Variablen *EF528* bis *EF537*); im Mikrozensus 2005 hingegen ist dieser Variablen-Typ in 33 Kategorien fein unterteilt (Variablen *EF669* bis *EF701*). Ein weiteres Beispiel für Variablen-Typen sind die neu eingeführten Variablen der Anzahl der ledigen Kinder in der Familie oder Lebensform, die in 33 Alterskategorien unterteilt sind (Variablen *EF770* bis *EF802*). Somit erleichtert der Mikrozensus 2005 als Instrumentarium spezifizierte Analysen. Die wohl bedeutendsten Neuerungen im Mikrozensus 2005 beziehen sich auf die Umstellung auf das neue Konzept der Lebensformen und die neue Abgrenzung bei Haushalten. Auf diesen Besonderheiten liegt im folgenden Kapitel der Fokus.

2.1 Konzepte und Definitionen

In diesem Kapitel wird zunächst die neue Definition der Bezugsperson des Haushaltes gemäß der Formulierung des Statistischen Bundesamtes genannt. Dabei ist die jeweilige Betrachtungsebene zu beachten. Der Mikrozensus ist als Mehrzweckstichprobe zu charakterisieren, d.h. die Informationen, die in der Datenerhebung gesammelt werden, sind auf mehreren Ebenen angesiedelt: Das sind

1. die Wohnungsebene
2. die Haushaltsebene
3. die Familien- und Lebensformebene
4. die Personenebene

Diese vier Ebenen weisen eine hierarchische Struktur auf.

Die Wohnung ist die hierarchisch oberste Analyseeinheit. Dies impliziert, dass sich in einer Wohnung mehrere Haushalte befinden können oder beide Ebenen deckungsgleich sind. Analog verhält es sich mit den weiteren Ebenen: Der Haushalt ist den beiden anderen Einheiten hierarchisch übergeordnet, da sich in einem Haushalt sowohl mehrere Lebensformen oder Familien als auch Personen, die derartigen Institutionen nicht angehören, befinden können.⁶

Die Analyseeinheit Wohnungsebene findet jedoch nur in den so genannten Wohnungssubstichproben des Mikrozensus Verwendung. Diese werden nur alle vier Jahre erhoben, das bedeutet, es liegen beispielsweise zu den Jahren 1998, 2002, und bald 2006 und 2010 Wohnungssubstichproben vor. In den anderen Mikrozensen ist dementsprechend der Haushalt die höchste Ebene. Im Zuge der Datenerhebung erfolgt bereits die Abgrenzung von Wohnungen (auf Basis der „systemfreien Nummer: laufende Wohnung im Auswahlbezirk EF7), Haushalten (mithilfe der „systemfreien Nummer: laufender Haushalt im Auswahlbezirk“ EF3) und Personen. Es handelt sich hierbei um „Erhebungseinheiten“. Im Anschluss werden im Rahmen der Datenaufbereitung „Auszähleinheiten“ gebildet. Im Nachhinein werden die „Auszähleinheiten“, je nachdem, ob das traditionelle Familienkonzept oder das Lebensformenkonzept greifen soll, durch die statistischen Landesämter, die Personen nach bestimmten Erhebungsmerkmalen zusammen fassen, gebildet. Dieser Prozess wird als „Familiensignierung“ bezeichnet (vgl. Nöthen, 2005: 27–28).

Je nach eigener Schwerpunktlage kommen unterschiedliche Betrachtungsebenen als homogene distinkte (über die Identifikatoren eindeutig abgrenzbare) Analyseeinheiten in Frage. Die Informationen über die Einheit wie beispielsweise der „Erwerbstyp des Haupteinkommensbeziehers im Haushalt EF736“ weisen dabei alle Individuen dieser Einheit, also in diesem Fall die Haushaltsmitglieder in Form der Ausprägung der abgeleiteten Variable EF736 auf.

Da diese Betrachtungsebenen prinzipiell verschiedene Zwecke erfüllen, dient beispielsweise die Haushaltsebene primär der Beantwortung von Fragen, die den Haushalt und die Wohngemeinschaft als ökonomische Mikrokosmen betreffen. Auf Familien- bzw. Lebensformenebene stehen dagegen die sozialen Beziehungen der Menschen untereinander im Mittelpunkt des Interesses (vgl. Nöthen, 2005: 27).

Anschließend werden im folgenden Unterkapitel die entsprechenden Definitionen der drei Analyseeinheiten Haushalt, Familie und Lebensform vorgestellt.⁷

Nach der Nennung der Definitionen wird dargestellt, mit welchen Identifikatoren bzw. Ordnungsnummern die Einheiten abgegrenzt werden können und wie die jeweiligen Bevölkerungsbegriffe als Filter bzw. zur Selektion zu verwenden sind.

⁶ Einen ausführlichen Überblick über die genannten Analyseeinheiten bieten Lengerer, Bohr und Janssen (2005).

⁷ Die im Mikrozensus 2005 verwendeten Definitionen können auch auf der Web-Seite des GML nachgelesen werden: http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_2005/Definitionen/mz05_def_index.htm

2.1.1 Haushalt

Der Haushalt wird folgendermaßen definiert (vgl. Statistisches Bundesamt. 2006a: 91):

„Als (Privat)-Haushalt zählt jede zusammen wohnende und eine wirtschaftliche Einheit bildende Personengemeinschaft (Mehrpersonenhaushalte) sowie Personen, die allein wohnen und wirtschaften (Einpersonenhaushalte, zum Beispiel auch Einzeluntermieter).“

Es werden dementsprechend nur soziale Kohabitationen erfasst, die als ökonomische Einheit nicht nur zusammen wohnen, sondern auch zusammen wirtschaften, d.h. beispielsweise kochen und den Lebensunterhalt finanzieren. Die deutsche amtliche Statistik richtet sich dabei nach dem von den Vereinten Nationen empfohlenen „housekeeping unit concept“ (vgl. Lengerer et al. 2005: 6).

Weiter heißt es in der Definition des Statistischen Bundesamtes (vgl. Statistisches Bundesamt. 2006a: 91):

„Zum Haushalt können verwandte und familienfremde Personen gehören (zum Beispiel Hauspersonal). Gemeinschaftsunterkünfte gelten nicht als Haushalte, können aber Privathaushalte beherbergen (zum Beispiel den Haushalt des Anstaltsleiters). Personen mit mehreren Wohnsitzen (Wohnungen am Haupt- und einem oder mehreren Nebenwohnsitzen) werden mehrfach gezählt (siehe [unten] Bevölkerung in Privathaushalten). In einem Haushalt können gleichzeitig mehrere Familien/Lebensformen (zum Beispiel ein Ehepaar ohne Kinder sowie eine allein erziehende Mutter mit Kindern) leben. Haushalte werden üblicherweise nicht nach Haupt- und Nebenwohnsitz unterschieden, da sie an jedem Wohnsitz Wohnraum und Infrastruktureinrichtungen in Anspruch nehmen. Stehen ökonomische Fragestellungen (Einkommen und Verbrauch) im Vordergrund, werden zur Vermeidung von Mehrfachzählungen ausschließlich Haushalte am Hauptwohnsitz betrachtet. Die Zuordnung der Haushalte nach Haupt- und Nebenwohnsitz erfolgte in den Mikrozensen bis einschließlich 2004 über den Wohnsitz der Haushaltsbezugsperson (siehe Bezugsperson des Haushalts) und ab dem Mikrozensus 2005 über den Haupteinkommensbezieher des Haushalts (siehe Haupteinkommensbezieher des Haushalts). Zu den in Privathaushalten am Hauptwohnsitz (Nebenwohnsitz) lebenden Personen zählen entsprechend alle Haushaltsmitglieder mit Bezugsperson beziehungsweise Haupteinkommensbezieher des Haushalts am Ort der Hauptwohnung (Nebenwohnung). Insofern kann der persönliche Wohnsitz einzelner Mitglieder von Mehrpersonenhaushalten vom Wohnsitz der Einheit „Haushalt“ abweichen.“

An dieser Stelle ist anzumerken, dass die Analyseeinheit „Haushalt“ in den Scientific Use Files mittels der Ordnungsnummern bzw. Identifikatoren Bundesland (EF1), Systemfreie Nr.: Auswahlbezirk (EF3) und Systemfreie Nr.: Haushalt im Auswahlbezirk (EF4) abgegrenzt werden kann.

In Kapitel 3 wird genauer erläutert, wie dies in STATA und SPSS bewerkstelligt werden kann und welche Logik zugrunde liegt.

Die Abgrenzung erfolgt ebenso wie schon im vorherigen Kapitel beschrieben über die Bevölkerung in Privathaushalten (EF31=1) und die Haushaltsbezugsperson (EF662>=1 & EF662<=4) unter Berücksichtigung der Einheiten.

2.1.2 Bezugsperson des Haushalts

Die Bezugsperson des Haushalts wird vom Statistischen Bundesamt (2006a: 88) wie folgt definiert:

„Um Haushalte statistisch auswerten und darstellen zu können, verwendet der Mikrozensus eine Bezugsperson des Haushalts (Haushaltsbezugsperson). Das ist die erste im Fragebogen eingetragene Person. Die Erhebungsmerkmale der Bezugsperson des Haushalts (zum Beispiel Alter, Geschlecht, Familienstand) wurden dann in den Mikrozensen bis einschließlich 2004 – stellvertretend für die gesamte Einheit „Haushalt“ – in der Statistik nachgewiesen. Personen unter 15 Jahren sind als Bezugsperson eines Haushalts ausgeschlossen. Die Reihenfolge im Erhebungsbogen ist: Ehegatten, Kinder, Verwandte, Familienfremde. Ab dem Mikrozensus 2005 wird die gesamte Einheit „Haushalt“ stellvertretend durch die Erhebungsmerkmale des Haupteinkommensbeziehers des Haushalts statistisch nachgewiesen.“

Beispielsweise könnte es interessant sein, zu untersuchen, ob, unter welchen Umständen und wie häufig der Mann in einer gemischtgeschlechtlichen Lebensgemeinschaft der Haupteinkommensbezieher ist und ob dies dem so genannten „male-breadwinner-model“ entspricht.

Als Abgrenzung verwendet man die Bevölkerung in Privathaushalten ($EF31=1$) und den Haupteinkommensbezieher ($EF662 \geq 1 \ \& \ EF662 \leq 4$).

Es ist jedoch darauf hinzuweisen, dass es im Mikrozensus 2005 auch ein Merkmal für die erste im Fragebogen eingetragene Person gibt, ohne dass der oder die HaupteinkommensbezieherIn berücksichtigt wird: die Variable Bezugsperson im Haushalt $EF37=1$.

Somit lassen sich für Vergleiche mit dem Mikrozensus 2004 oder noch früheren Datensätzen eigene abgeleitete Variablen bilden, die über die frühere Form der Haushaltsbezugsperson abgegrenzt werden.

2.1.3 Lebensform

Die Analyseeinheit Lebensform wird durch das Lebensformenkonzept definiert, das in der amtlichen Statistik 1996 eingeführt wurde. Kriterien zur Bestimmung von Lebensformen sind Partnerschaft und Elternschaft. Aus der Kombination der beiden Merkmale ergeben sich folglich vier Typen von sozialen Lebensformen:

Paare mit Kindern, Paare ohne Kinder, Alleinerziehende und Alleinstehende. Es werden sowohl gleichgeschlechtliche als auch gemischtgeschlechtliche Paarbeziehungen erfasst.

Problematisch bei Alleinerziehenden oder Alleinstehenden ist jedoch, dass nur soziale Beziehungen innerhalb der Grenzen eines Haushaltes erfasst werden und somit eventuelle vorhandene Beziehungen zu Partner(n) eines anderen Haushaltes nicht abgebildet werden.

Das Statistische Bundesamt (2006a: 92) erläutert das „Lebensformenkonzept“, unter Herausstellung der Besonderheiten des Mikrozensus 2005, folgendermaßen:

„Seit 1996 wird im Mikrozensus die Frage nach eine(m)/r Lebenspartner/in im Haushalt gestellt, deren Beantwortung freiwillig ist. Zwischen 1996 und 2004 richtete sie sich an alle nicht mit der Haushaltsbezugsperson (siehe Bezugsperson des Haushalts) verwandten oder verschwägerten Haushaltsmitglieder und lautete: „Sind Sie Lebenspartner(in) der ersten Person?“. 2005 wurde im Mikrozensus erstmals allen mindestens 16-jährigen Haushaltsmitgliedern ohne Ehepartner/in im Haushalt die Frage nach einer Lebenspartnerschaft gestellt. Sie hieß: „Sind Sie Lebenspartner, Lebenspartnerin einer Person dieses Haushalts?“. Damit können ab dem Mikrozensus 2005 erstmals mehrere Lebensgemeinschaften in einem Haushalt erhoben werden. In den Mikrozensusen 1996 bis 2004 konnte dagegen jeder Haushalt höchstens eine Lebensgemeinschaft angeben, da die entsprechende Frage ausschließlich auf eine Lebenspartnerschaft mit der Bezugsperson des Haushalts abstellte. Die Frage nach einer Lebenspartnerschaft im Haushalt ist neutral formuliert und lässt bewusst das Geschlecht der Befragten außer Betracht. Damit können auch gleichgeschlechtliche Paare ihre Lebensgemeinschaft angeben. Unerheblich ist, ob die Partnerschaft als eingetragene Lebenspartnerschaft nach dem im Jahr 2001 eingeführten Lebenspartnerschaftsgesetz (LpartG) registriert wurde. Die seit dem Mikrozensus 1996 erhobenen Informationen zu Lebenspartnerschaften ermöglichen das so genannte Lebensformenkonzept. Es ist ab dem Berichtsjahr 2005 der Standard für die Veröffentlichung familienbezogener Ergebnisse aus dem Mikrozensus. Inhaltlich berücksichtigt das Lebensformenkonzept – neben den traditionellen Lebensformen, wie Ehepaaren – insbesondere alternative Lebensformen, wie nichteheliche (gemischtgeschlechtliche) und gleichgeschlechtliche Lebensgemeinschaften oder Alleinstehende. Grundlage für die Darstellung von Ergebnissen nach dem Lebensformenkonzept ist die so genannte „Bevölkerung (Lebensformenkonzept)“.

Die Einheit Lebensform wird wiederum mithilfe der Ordnungsnummern Bundesland (EF1), Systemfreie Nr.: Auswahlbezirk (EF3) und Systemfreie Nr.: Haushalt im Auswahlbezirk (EF4), sowie zusätzlich mit der Variablen Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept) (EF27) abgegrenzt.

2.1.4. Familie

Die Analyseeinheit Familie ist ebenso wie die der Lebensform dem Haushalt hierarchisch untergeordnet. Der Familienbegriff im Mikrozensus 2005 muss zunächst vom traditionellen Familienbegriff abgegrenzt werden: Mit dem Mikrozensus 2005 vollzog sich in der amtlichen Statistik ein Wechsel von einem traditionellen Familienbegriff hin zu einem kindzentrierten. Während bis zum Mikrozensus 2004 Familie über das Ehe- und Abstammungsprinzip definiert wird, d.h. zur Klassifikation als Familie ist es ausreichend, wenn entweder eine Ehe oder eine Eltern-Kind-Gemeinschaft vorliegt, liegt im Mikrozensus 2005 eine Familie nur dann vor, wenn ledige Kinder vorhanden sind.

Die Analyseeinheit „Familie“ wird vom Statistischen Bundesamt (2006a: 89) folgendermaßen umrissen:

„Die Familie im statistischen Sinn umfasst im Mikrozensus – abweichend von früheren Veröffentlichungen – alle Eltern-Kind-Gemeinschaften, das heißt Ehepaare, nichteheliche (gemischtgeschlechtliche) und gleichgeschlechtliche Lebensgemeinschaften sowie allein erziehende Mütter und Väter mit ledigen Kindern im Haushalt. Einbezogen sind in diesen Familienbegriff – neben leiblichen Kindern – auch Stief-, Pflege- und Adoptivkinder ohne Altersbegrenzung. Damit besteht eine statistische Familie immer aus zwei Generationen (Zwei-Generationen-Regel): Eltern/-teile und im Haushalt lebende ledige Kinder. Kinder, die noch gemeinsam mit den Eltern in einem Haushalt leben, dort aber bereits eigene Kinder versorgen, sowie Kinder, die nicht mehr ledig sind oder mit eine(m)/r Partner/in in einer Lebensgemeinschaft leben, werden im Mikrozensus nicht der

Herkunftsfamilie zugerechnet, sondern zählen statistisch als eigene Familie beziehungsweise Lebensform [...]”⁸

Die Einheit „Familie“ kann im Scientific Use File Mikrozensus 2005 mittels der Ordnungsnummern Bundesland (EF1), Systemfreie Nr.: Auswahlbezirk (EF3) und Systemfreie Nr.: Haushalt im Auswahlbezirk (EF4) und Laufende Nr.: Familie im Haushalt "traditionelles Familienkonzept" (EF25) bzw. Laufende Nr.: Familie/Lebensform im Haushalt "Lebensformenkonzept" (EF27) abgegrenzt werden.

Im Mikrozensus 1996 bis 2004 stehen für die Abgrenzung der Einheit Familie ebenfalls die Variablen EF1, EF3, EF4 und Laufende Nr.: Familie im Haushalt "traditionelles Familienkonzept" (EF28) bzw. Lfd. Nummer der Lebensgemeinschaft im Haushalt „Lebensformenkonzept“ (EF643) zur Verfügung. Die Abgrenzung im Mikrozensus 2005 erfolgt über die am Hauptwohnsitz befindliche Bezugsperson der Familie/Lebensform (EF761>=1 & EF761<=2) und der Stellung als Bezugsperson innerhalb der Familie/Lebensform (EF762=1) sowie das Kriterium, ob es ledige Kinder in der Lebensgemeinschaft gibt (EF809>=1 & EF809<=4).

2.1.5 Abgrenzungen der Bevölkerungsbegriffe

Im Folgenden werden anhand der Tabelle 2 die im Mikrozensus 2005 verwendeten Bevölkerungsbegriffe und Abgrenzungen denselbigen des Mikrozensus 2004 gegenüber gestellt.

Unterschiede ergeben sich vor Allem durch das neue Konzept der Haushaltsbezugsperson (siehe Privathaushalte) und den kindzentrierten Familienbegriff im Mikrozensus 2005 (siehe Familien nach dem Lebensformenkonzept).

Weiterhin ist darauf zu achten, welches Bevölkerungskonzept zugrunde liegt. Im Mikrozensus entspricht die Zielpopulation („*target population*“) der Erhebungspopulation („*survey population*“); respektive wird die gesamte wohnberechtigte Bevölkerung in Deutschland samt Personen in Gemeinschaftsunterkünften (wie Gefängnisse, Heime, Krankenhäuser) bis auf wenige Ausfälle erfasst. Bei Auswertungen auf der Ebene des Haushaltes wird jedoch nur die Bevölkerung in Privathaushalten betrachtet.

Zur Grundgesamtheit zählen laut Statistischem Bundesamt (2003: 15):

„alle in der Gemeinde wohnhaften Personen, unabhängig davon, ob sie noch eine weitere Wohnung oder Unterkunft besitzen und von wo aus sie zur Arbeit oder Ausbildung gehen, bzw. wo sie sich überwiegend aufhalten“.

Oftmals wird daher nur die Bevölkerung am Ort der alleinigen bzw. Hauptwohnung betrachtet, um Doppelzählungen zu vermeiden: Es wäre theoretisch möglich, dass eine Person sowohl an ihrem Hauptwohnsitz als auch an ihrem Nebenwohnsitz befragt wird oder dass eine Person beispielsweise

⁸ Im Aufsatz von Duschek und Wirth (2005) findet dieser neue Familienbegriff Verwendung.

an ihrem Nebenwohnsitz befragt wird und am Hauptwohnsitz im Rahmen der Selbstauskunft nochmals befragt wird. Dementsprechend wäre diese Person zweimal im Mikrozensus erfasst. Die Erläuterung der Variablen der folgenden Tabelle 2 befinden sich im Anhang.

Tabelle 2: Bevölkerungsbegriffe und Abgrenzungen im Mikrozensus SUF 2005

Bevölkerungsbegriff Subpopulation	Ab MZ 2005 verwendete Filter zur Abgrenzung der Subpopulation	Bis einschließlich MZ 2004 verwendete Filter zur Abgren- zung der Subpopulation
Bevölkerung am Haupt- und Nebenwohnsitz (früher: wohnberechtigte Bevölkerung)	EF30>=1 & EF30<=3	EF505>=1 & EF505<=3
Privathaushalte	EF31=1 & EF662>=1 & EF662<=4	EF506=1 & EF507=1
Bevölkerung in Privathaushalten	EF31=1	EF506=1
Bevölkerung am Ort der alleinigen beziehungsweise Hauptwohnung (kurz: Bevölkerung)	EF30>=1 & EF30<=2	EF505=1 EF505=2
Familien (Lebensformenkonzept) ^a	EF761>=1 & EF761<=2 & EF762=2 & EF809>=1 & EF809<=4	EF513=1 & EF514=1
Familien (trad. Familienkonzept) ^b	EF863=1 & EF864=1	EF508=1 & EF509=1
Bevölkerung am Familienwohnsitz (traditionelles Familienkonzept)	EF863==1	EF508=1
Lebensgemeinschaften (Lebensformenkonzept)	EF761>=1 & EF761<=2 & (EF763=3 EF763=8)	EF513=1 & EF514=1
Bevölkerung in Familien/Lebensformen am Hauptwohnsitz (Lebensformenkonzept)	EF761>=1 & EF761<=2	EF513=1
Erwerbstätige	EF29=1	(EF505=1 EF505=2) & EF504=1

Quellen:

http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_2005/hochrechnungsfaktoren_mz05.htm und
http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_2004/hochrechnungsfaktoren_mz04.htm

3 Rekonstruktion von abgeleiteten Variablen

Ziel des vorliegenden Kapitels ist es, anhand von acht Variablen beispielhaft aufzuzeigen, wie abgeleitete Variablen selbst generiert werden können.⁹

Vor Beginn der eigenen Rekonstruktion muss festgestellt werden, auf welcher Ebene (Haushalts-, Familien- oder Lebensgemeinschaftsebene) Kontextinformationen benötigt werden.

⁹ Einen Überblick darüber, wie abgeleitete Variablen mithilfe verschiedener Befehlssequenzen in STATA und SPSS gebildet werden können, zeigt der GESIS-ZUMA-Methodenbericht von John (2007). Auch der Methodenbericht von Lengerer und Boehle (2006) erläutert ausführlich, wie abgeleitete Variablen in SPSS erzeugt werden können.

Zudem besteht die Option, entweder Informationen über eine ganze Einheit (beispielsweise den Haushalt) den Personen einer Einheit (z.B. Zahl der Erwerbstätigen im Haushalt *EF664*) oder Informationen über einer sich in einer Einheit befindenden Person (z.B. den Erwerbstyp des Haupteinkommensbeziehers *EF736*) den anderen Personen dieser Einheit zuzuspielen.

Die Analyseeinheiten müssen zunächst identifiziert werden. Dies ist über die hierarchische Struktur der Daten möglich: Beispielsweise ist jeder Haushalt über eine charakteristische Merkmalskombination des Bundeslandes *EF1*, der systemfreien Nummer des Auswahlbezirks *EF3* und der systemfreien Nummer des Haushalts im Auswahlbezirk *EF4* gekennzeichnet.

Die Daten müssen vor der Analyse nach diesen drei Ordnungsnummern sortiert werden.

Dies geschieht mithilfe des „*sort*“-Befehls in STATA:

```
sort EF1 EF3 EF4
```

bzw. des „*sort-cases*“-Befehls in SPSS:

```
sort cases by EF1 EF3 EF4.
```

Die vollständige STATA- und SPSS-Syntax kann im Anhang nachvollzogen werden.

Bei der Sortierung müssen die Variablen in exakt der genannten Reihenfolge aufgelistet werden, da ihnen wie beschrieben eine hierarchische Struktur zugrunde liegt. Das bedeutet, dass zuerst alle Fälle nach den verschiedenen Bundesländern sortiert werden, in den jeweiligen Bundesländern erhalten sie jeweils eine Ausprägung für den Auswahlbezirk in diesem Bundesland und in dem entsprechenden Auswahlbezirk je eine Nummer für den entsprechenden Haushalt.

Betrachtet man nun jedoch nicht nur die Haushaltsebene, sondern auch die untergeordnete Familien- oder Lebensformebene¹⁰, müssen die Analyseeinheiten bei der Familie im traditionellen Sinn zusätzlich nach dem Identifikator der laufenden Nummer der Familie im Haushalt *EF25*

sortiert werden. Im Folgenden wird dies anhand der Rekonstruktion der Variable der Zahl der ledigen Kinder insgesamt in der Familie *EF866* veranschaulicht.

Möchte man hingegen Aussagen über eine Lebensgemeinschaft treffen, sortiert man die Variablen nicht nur nach *EF1*, *EF3* und *EF4*, sondern zusätzlich auch nach der laufenden Nummer der Familie oder Lebensform im Haushalt *EF27*.

Illustriert wird diese Vorgehensweise anhand der Variablen

- Zahl der nichtehelichen Lebensgemeinschaften im Haushalt *EF713* (Beispiel 2),
- Zahl der ledigen Kinder in der Familie/Lebensform insgesamt (ohne Altersbegrenzung) *EF770* (Beispiel 5), Erwerbstyp der Bezugsperson in der Familie/Lebensform *EF824* (Beispiel 6) und

¹⁰ Familien und Lebensformen sind dem Haushalt hierarchisch untergeordnet, da in einem Haushalt mehrere Familien oder Lebensformen existieren können, vgl. Kapitel 2.1.2.

- Erwerbstyp des Ehe- oder Lebenspartners *EF848* (Beispiel 7).

Um einen Eindruck über das Analysepotenzial zu vermitteln, das sich durch die Kenntnis ergibt, wie man selbst abgeleitete Variablen bilden kann, werden in den nächsten Unterkapiteln acht im Mikrozensus 2005 vorhandene Variablen anhand der STATA-Syntax beschrieben.

Falls Interesse an der „Übersetzung“ in SPSS besteht, sei auf den Anhang verwiesen.

Alle der beschriebenen Variablen lassen sich am geschicktesten mithilfe des Befehls *„extended generate“* (abgekürzt *„egen“*) bilden.¹¹ Die Eleganz des Befehls drückt sich darin aus, dass die Aggregation der Kontextinformation über die Analyseeinheit (also Aggregatinformationen auf der Haushalts-, Familien- oder Lebensformenebene) und die Zuweisung dieser Information zu den der Einheit zugehörigen Individuen als auch die Sortierung nach den Identifikatoren in einem Schritt erfolgt. Daher ist in diesem Fall keine vorherige Sortierung nach den Identifikatoren, bzw. Ordnungsnummern notwendig.

3.1 Replikation der Variable *EF664* „Zahl der Erwerbstätigen im Haushalt“

Da es möglich ist, dass aufgrund unplausibler Daten im Zuge der Dateneingabe oder -erfassung mehr als eine Bezugsperson im Haushalt vorhanden sind, wird zunächst mithilfe des *„egen“*-Befehls die Anzahl der Bezugspersonen im Haushalt gebildet. Im zweiten Schritt wird nur für diejenigen Haushalte mit einer Bezugsperson im Haushalt die abgeleitete Variable der Zahl der Erwerbstätigen im Haushalt *EF664* gebildet, andernfalls werden die Personen des Haushaltes auf *„Missing“* gesetzt. Dies vollzieht sich mithilfe einer *„else“*-Schleife in einem einzigen Befehl (s. unten).

Wie im Folgenden zu sehen, gleicht die Befehlsstruktur von *„egen“* der von *„generate“* (*„gen“*). Jedoch beinhaltet sie eine große und ständig wachsende Zahl von Erweiterungen, d.h. eigentlich eine Reihe von hintereinander ausgeführten, mehr oder weniger komplizierten *„generate“*- und *„replace“*-Kommandos und lässt sich für diverse Operationen verwenden (vgl. Kohler und Kreuter 2006: 94).

Die Anzahl der Bezugspersonen wird erzeugt, indem nach dem Befehl *„egen“* der Name der Variable, die erzeugt werden soll (hier: *n_bezper_hh* für die Anzahl *n* der Bezugspersonen im Haushalt), folgt, dann ein Gleichheitszeichen und schließlich die *„egen“*-Funktion. Letztere zeichnet sich dadurch aus, dass innerhalb der Klammern die benötigte Abgrenzung auf Individualebene steht (also die Bevölkerung in Privathaushalten [*EF31==1*] und die Haushaltsbezugsperson [*EF662>=1 & EF662<=4*]). Die der Klammer vorstehende Funktion *„total“* bewirkt, dass die Summe der Bezugspersonen im Haushalt gebildet wird. Mithilfe des unverzichtbaren Zusatzes nach dem Komma, *„by (EF1 EF3 EF4)“*

¹¹ Weitere mögliche Befehle sind *„merge“* und *„joinby“*. Ihre Anwendung wird in John (2007) dargestellt.

erfolgt programm-intern die genannte Sortierung der Daten und damit die Identifikation der Haushalte:

```
egen n_bezper_hh = total(EF31==1 & EF662>=1 & EF662<=4), by(EF1 EF3 EF4)
```

In einem zweiten Schritt kann nun, auch unter Zuhilfenahme des *egen*-Befehls, die Variable der Anzahl der Erwerbstätigen im Haushalt *EF664* gebildet werden. Die Filtersetzung vollzieht sich innerhalb der Klammer; es werden nur Erwerbstätige (*EF29=1*) in Privathaushalten (*EF31=1*) betrachtet:

```
if n_bezper_hh==1 {  
    egen v664 = total(EF29==1 & EF31==1), by(EF1 EF3 EF4)  
}  
else {  
    replace v664==.  
}
```

Danach sollte kontrolliert werden, ob die Rekonstruktion der abgeleiteten Variable *EF664*, also *v664*, korrekt gebildet wurde. Dies kann man mithilfe einer Kreuztabellierung simpel prüfen: Die Zuweisung war erfolgreich, wenn sich alle Werte auf der Diagonalen der Tabelle befinden.

```
tab v664 EF664 if EF31==1, miss
```

Leider liegen nicht alle Fälle auf der Hauptdiagonalen. Ein Blick auf die Merkmalsverteilung der ersten 25 Haushalte zeigt, dass die eigene Rekonstruktion der Variable *v664* offensichtlich korrekt gebildet wurde, da die Anzahl der Fälle in einem Haushalt, bei denen der Erwerbsstatus der Person als erwerbstätig klassifiziert wird (*EF29=1*), korrekt gebildet wurde.

```
list EF1 EF3 EF4 EF29 EF31 EF664 v664 in 1/25, nolab sepby(EF1 EF3 EF4)
```

Nach dem Zusammenfassen der Zahl der Erwerbstätigen im Haushalt bei sechs Personen und mehr zu der Kategorie "mindestens 6 Personen" durch folgende Syntax ergibt sich Tabelle 4:

```
recode v664 (6/9=6), gen(y664) copyrest  
tab EF664 y664, miss
```

Tabelle 4: Kontrolle der Hauptdiagonalen durch Kreuztabellierung der abgeleiteten Variable der Anzahl der Erwerbstätigen im Haushalt *EF664* und ihrer Rekonstruktion *v664*

Erwerbstätige im Haushalt: Anzahl <i>EF664</i>	Replikation Zahl der Erwerbstätigen im Haushalt <i>v664</i>							Total
	0	1	2	3	4	5	6	
0	138.721	3	0	0	0	0	0	138.724
1	0	148.317	2	1	0	0	0	148.320
2	0	6	146.032	0	0	0	0	146.038
3	0	0	0	30.377	0	0	0	30.377
4	0	0	0	0	7.500	0	0	7.500
5	0	0	0	0	0	905	0	905
6 und mehr	0	0	0	0	0	0	158	158
Gemeinschaftsunterkünfte	5.217	0	0	0	0	0	0	5.217
Total	143.938	148.326	146.034	30.378	7.500	905	158	477.239

Es liegen offensichtlich Diskrepanzen im Mikrozensus 2005 vor. Abweichungen sind in der Tabelle hervorgehoben. Diese werden in Kapitel 4 ausführlich behandelt.

3.2 Replikation der Variable *EF713* „Zahl der nichtehelichen Lebensgemeinschaften im Haushalt“

Für die Rekonstruktion der Zahl der nichtehelichen Lebensgemeinschaften im Haushalt wird zusätzlich als Identifikator die laufende Nummer der Familie oder Lebensform im Haushalt nach dem Lebensformenkonzept *EF27* benötigt. Die Information über die jeweilige Anzahl wird jedem Mitglied der Lebensgemeinschaften innerhalb jedes Haushaltes zugespielt. Hierbei wird analog zum ersten Beispiel zunächst die Anzahl der Bezugspersonen innerhalb der Lebensform ermittelt werden (Stellung innerhalb der Familie/Lebensform als Bezugsperson *EF35=1*). Sie sollte auch nicht größer als Eins sein. Es erfolgt wiederum eine Beschränkung auf Privathaushalte (*EF31=1*):

```
egen n_bezper_hh_lebensform = total(EF31==1 & EF35==1), by(EF1 EF3 EF4 EF27)
```

Bei der eigentlichen Rekonstruktion der abgeleiteten Variable Zahl der Lebensformen im Haushalt *EF713* erfolgt zudem eine Beschränkung auf den Familien- oder Lebensformtyp *EF809* der Ehepaare mit ledigen Kindern (*EF809=1*) und die nichteheliche Lebensgemeinschaft mit ledigen Kindern (*EF809=6*).

```
if n_bezper_hh_lebensform==1 {
```

```
egen v713 = total(EF31==1 & EF35==1 & (EF809==2 | EF809==6)), by(EF1 EF3 EF4)
```

```
}
```

```
else {
```

```

        replace v713==.
    }

```

Auch hier weisen die Kreuztabellierung als auch der „list“-Befehl auf einige inkongruente Daten hin (vgl. Kapitel 4).

3.3 Replikation der Variable *EF736* „Erwerbstyp: Haupteinkommensbezieher“

Bei dieser Form der Rekonstruktion werden Informationen über eine Person der Analyseeinheit, den Haupteinkommensbezieher, auf alle anderen Mitglieder der Einheit, die Haushaltsmitglieder, übertragen. Die Befehlsstruktur unterscheidet sich etwas von der der vorherigen Beispiele:

Statt „total“ wird der Term „max“ vor den in der Klammer stehenden Abgrenzungen gewählt. Zudem wird die interessierende Variable, bei der nicht nur eine bestimmte (wie *EF29=1* im ersten Beispiel, s. Kapitel 3.1), sondern alle vorhandenen Merkmalsausprägungen interessieren. In diesem Fall ist demnach die Frage, welche Ausprägung der Haupteinkommensbezieher bei der Variable Erwerbstyp *EF29* aufweist. Möglich wären die Ausprägungen 1 für "Erwerbstätige", 2 für "Erwerbslose", 3 für "Arbeitssuchende Nichterwerbspersonen", 4 für "Sonstige Nichterwerbspersonen" und 0 bei Item-Nonresponse. Der Befehl „max“ führt nun dazu, dass die Ausprägung des Erwerbstyps des Haupteinkommensbeziehers (*EF662=1, 2, 3 oder 4*) den Personen des Haushaltes in Privathaushalten zugespielt wird.

```

if n_bezper_hh==1 {
    egen v736 = max( EF29 * (EF31==1 & (EF662==1 | EF662==2 | EF662==3 | EF662==4))), by(EF1
    EF3 EF4)
}
else {
    replace v736==.
}

```

Auch hierbei weist die Kontrolle auf Inkonsistenzen des Mikrozensus 2005 hin.

3.4 Replikation der Variable *EF758* „Erwerbstyp: Haushaltsbezugsperson“

Analog zum vorherigen Beispiel wird nun auch unter Verwendung des „*max*“-Befehls der Erwerbstyp der Haushaltsbezugsperson gebildet. Da im Mikrozensus 2005 die Haushaltsbezugsperson dem Haupteinkommensbezieher entspricht, sollte die folgende abgeleitete Variable mit derjenigen des dritten Beispiels (*EF736*) übereinstimmen. Die Befehlsstruktur lautet:

```
if n_bezper_hh==1 {  
    egen v758 = max( EF29 * (EF31==1 & EF661==1)), by(EF1 EF3 EF4)  
}  
else {  
    replace v758==.  
}
```

Mithilfe der Einschränkung auf die Haushaltsbezugsperson (*EF661=1*) wird diese Variable jedoch einfacher als die im Kapitel 3.3 beschriebene gebildet. Doch auch hierbei treten bei 60 Haushalten Inkongruenzen auf.

3.5 Replikation der Variable *EF770* „Zahl der ledigen Kinder in der Familie/Lebensform insgesamt (ohne Altersbegrenzung)“

Bei der Replikation der Variablen der Zahl der ledigen Kinder in der Familie oder Lebensform erfolgt neben der Beschränkung auf Privathaushalte (*EF31=1*) eine Abgrenzung der Stellung innerhalb der Familie oder Lebensform auf die ledigen Kinder in der Familie (*EF35=3*). Die Sortierung der Ordnungsnummern vollzieht sich zusätzlich zu *EF1*, *EF3* und *EF4* auch über die laufende Nummer der Familie oder Lebensform im Haushalt (*EF27*).

```
if n_bezper_hh_lebensform==1 {  
    egen v770 = total(EF31==1 & EF35==3),  
    by(EF1 EF3 EF4 EF27)  
}  
else {  
    replace v770==.  
}
```

Die Definition als lediges Kind erfolgt dabei ohne Altersbegrenzung. Doch auch diesmal sind Abweichungen im Mikrozensus 2005 zu konstatieren.

3.6 Replikation der Variable *EF824* „Erwerbstyp der Bezugsperson in der Familie/Lebensform“

Auch bei der Variable Erwerbstyp der Bezugsperson in der Familie oder Lebensform (*EF824*) muss der Datensatz ebenso wie in Kapitel 3.5 mithilfe der Identifikatoren *EF1*, *EF3* und *EF4* und der laufenden Nummer der Familie oder Lebensform im Haushalt (*EF27*) gebildet werden. Der Erwerbstyp der Bezugsperson (*EF35=1*) wird wie auch in den Kapitel 3.3 und 3.4 mithilfe des „max“-Befehls allen Mitgliedern der Einheit, hier jedoch Mitgliedern der Familie oder Lebensform und nicht denen des Haushaltes zugewiesen. Natürlich können die Einheiten Haushalt und Familie oder Lebensform deckungsgleich sein, sie müssen es aber nicht, da sich in einem Haushalt mehrere Familien oder Lebensformen befinden können.

```
if n_bezper_hh_lebensform==1 {  
    egen v824 = max( EF29 * (EF31==1 & EF35==1)), by(EF1 EF3 EF4 EF27)  
}  
else {  
    replace v824==.  
}
```

Wiederum bestätigt die Kontrolle mithilfe des tab- als auch list-Kommandos, dass es Abweichungen gibt, während unsere eigenen abgeleiteten Variablen mit den Daten übereinstimmen.

3.7 Replikation der Variable *EF848* „Erwerbstyp des Ehe- oder Lebenspartners“

Auch bei diesem Beispiel werden die Daten während des „egen“-Befehls nach den Ordnungsnummern *EF1*, *EF3* und *EF4* und der laufenden Nummer der Familie oder Lebensform im Haushalt (*EF27*) sortiert, da das Lebensformenkonzept greift. Da die Information über eine Person innerhalb der Lebensform, sei es eine eheliche oder uneheliche, auf die anderen Mitglieder derselben transferiert wird, muss, wie auch in den Kapiteln 3.3, 3.4 und 3.6 der „max“- Befehl verwendet werden. Die Abgrenzung innerhalb der Klammer bezieht sich wiederum auf die Bevölkerung in Privathaushalten (*EF31=1*) und zudem auf die Stellung innerhalb der Familie oder Lebensform als "Ehefrau oder Lebenspartner(in) der Bezugsperson in der Familie/Lebensform" (*EF35 =2*).

```
if n_bezper_hh_lebensform==1 {
```

```
egen v848 = max( EF29 * (EF31==1 & EF35==2)), by(EF1 EF3 EF4 EF27)
```

```
}
```

```
else {
```

```
    replace v848==.
```

```
}
```

Auch in diesem Fall weist die Kontrolle auf abweichende Daten hin.

3.8 Replikation der Variable *EF866* „Zahl der ledigen Kinder insgesamt in der Familie“

Die Rekonstruktion der abgeleiteten Variable der Zahl der ledigen Kinder insgesamt in der Familie (*EF866*) unterscheidet sich von den zuvor genannten Beispielen vor Allem dadurch, dass zusätzlich zu den dringend notwendigen Identifikatoren *EF1*, *EF3* und *EF4* außerdem die laufende Nummer der Familie im Haushalt (*EF25*) im Rahmen des traditionellen Familienkonzepts benötigt wird.

Im ersten Schritt werden wiederum unter Zuhilfenahme einer „*else*“-Schleife nur diejenigen Familien betrachtet, in denen es nur eine Bezugsperson gibt (*EF35=1*).

```
egen n_bezper_hh_trad = total(EF31==1 & EF35==1), by(EF1 EF3 EF4 EF25)
```

Im zweiten Schritt wird während der Aggregation und Zuweisung durch den „*egen*“-Befehl auch eine Einschränkung auf die Bevölkerung in Privathaushalten (*EF31=1*) vorgenommen. Darüber hinaus wird die Anzahl der ledigen Kinder insgesamt dadurch gebildet, dass Personen in der Stellung als lediges Kind innerhalb der Familie nach dem traditionellen Familienkonzept oder der Stellung als lediges Kind innerhalb der Familie oder Lebensform nach dem Lebensformenkonzept betrachtet werden.

```
if n_bezper_hh_trad==1 {
```

```
    egen v866 = total(EF31==1 & EF34==3 | EF35==3), by(EF1 EF3 EF4 EF25)
```

```
}
```

```
else {
```

```
    replace v866==.
```

```
}
```

Auch hier ergeben sich Abweichungen, die im kommenden Kapitel im Fokus stehen.

4 Diskrepanzen im Mikrozensus 2005

Bei der Erstellung der eigenen abgeleiteten Variablen mithilfe des Mikrozensus 2005 ist darauf zu achten, dass dieser einige Inkongruenzen enthält, die jedoch inhaltlich kaum Auswirkungen auf die durchzuführenden Analysen haben sollten, da sie nur im geringen Umfang auftreten.

Bei der Rekonstruktion der Variablen *EF664* „Anzahl der Erwerbstätigen im Haushalt“ sind in unserem Beispiel 17 Fälle, bzw. acht Haushalte inkongruent bezüglich der Ausprägung der Ursprungsvariable *EF664*. In zwei Haushalten ist zudem keine Haushaltsbezugsperson eingetragen.

Dabei wurden entweder nicht allen Personen eines Haushaltes dieselbe Information, in diesem Fall die Anzahl von Erwerbstätigen im Haushalt, zugespielt oder die Anzahl der Erwerbstätigen stimmt einfach nicht, bzw. in drei Haushalten ist auch beides der Fall (vgl. Kapitel 3.1).

Im zweiten Beispiel treten die größten Abweichungen auf: Bei der Rekonstruktion der Variablen *EF713* „Anzahl der nichtehelichen Lebensgemeinschaften im Haushalt“ stimmt die Ausprägung der Originalvariable *EF713* bei 1275 Fällen nicht mit der replizierten Variable überein. Dies scheint auf den ersten Blick vielleicht eine große Anzahl von Personen zu sein, im Verhältnis zur Substichprobe von $N = 477\,239$ sind es jedoch nur 0,267% der Fälle. Diese abweichenden Fälle verteilen sich auf 314 Haushalte (vgl. Kapitel 3.2).

Nach Überprüfung und Plausibilitätskriterien ergeben sich etwas geringere Inkongruenzen bei der Nachbildung der Variablen *EF736* „Erwerbstyp: Haupteinkommensbezieher“: Insgesamt weisen 108 Fälle abweichende Angaben der *EF736* auf, weil der Haupteinkommensbezieher zu den sonstigen Nichterwerbspersonen gezählt wurde, obwohl er eigentlich als erwerbslos zu klassifizieren ist. Daraus ergeben sich 50 Haushalte mit Aggregatangaben über den Erwerbstyp des Haupteinkommensbeziehers, die nicht mit den Daten übereinstimmen. In zwei Haushalten mit jeweils zwei ansässigen Personen ist jedoch gar keine Person als Haupteinkommensbezieher angegeben, sodass bei diesen mithilfe des „*egen*“-Befehls keine Bandsatzerweiterung gebildet werden konnte (vgl. Kapitel 3.3).

Ähnliche Zuordnungsprobleme zu den Erwerbslosen, mit der Variablen *EF29* „Erwerbstyp“ zusammenhängend, ergeben sich für die Replikation der Variablen *EF758* „Erwerbstyp: Haushaltsbezugsperson“, wiederum ist in denselben zwei genannten Haushalten keine Haushaltsbezugsperson vorhanden. Insgesamt sind sogar 137 Fälle, bzw. 60 Haushalte problematisch (vgl. Kapitel 3.4).

Bezüglich der eigenen Erstellung der Variablen *EF770* „Zahl der ledigen Kinder in der Familie/Lebensform insgesamt (ohne Altersbegrenzung)“ sind 117 Fälle und 96 Familien bzw. Lebensformen abweichend (vgl. Kapitel 3.5).

Ähnlich viele Inkongruenzen ergeben sich beim Vergleich der Variable *EF824* „Erwerbstyp der Bezugsperson in der Familie oder Lebensform“ und ihrer Rekonstruktion: 144 Personen stimmen mit

ihren Ausprägungen nicht überein. Die korrespondierenden abweichenden Haushalte belaufen sich auf die Anzahl von 62. Wird der kindzentrierte Familienbegriff verwendet, sind zudem 47 abweichende Familien bzw. Lebensformen identifizierbar (vgl. Kapitel 3.6).

Etwas weniger Abweichungen sind bei der Rekonstruktion der Variable *EF848* „Erwerbstyp des Ehe- oder Lebenspartners“ zu konstatieren: 67 Fälle weichen ab und verteilen sich auf 13 Haushalte, bzw. 13 Familien oder Lebensformen mit ledigem Kind (vgl. Kapitel 3.7).

Bei der „Zahl der ledigen Kinder insgesamt in der Familie“, *EF866*, weichen 119 Fälle ab. Wie die Datenanalyse zeigt, sind 12 Haushalte und 97 Familien von den Inkongruenzen betroffen (vgl. Kapitel 3.8).

Im Anhang ist auch ersichtlich, wie die abweichenden Personen und Haushalte identifiziert und ihre Anzahl bestimmt werden kann.

Zusammenfassend ist zu konstatieren, dass dieser Methodenbericht zeigen konnte, wie abgeleitete Variablen selbst konstruiert werden können. Eine Übertragung der beschriebenen Beispiele auf andere Themen sollte der forschenden Person somit möglich sein. Außerdem liefert der Bericht Hinweise darauf, wie die Daten des Scientific Use Files nach Plausibilitätskriterien auf eigene Weise bereinigt werden können.

Literatur

- Duschek, Klaus-Jürgen; Wirth, Heike (2005). Kinderlosigkeit von Frauen im Spiegel des Mikrozensus. Eine Kohortenanalyse der Mikrozensen 1987 bis 2003. In: *Wirtschaft und Statistik* 8/2005: 800-820.
- John, Kristina (2007). Rekonstruktion von abgeleiteten Variablen mittels zeilen- bzw. satzübergreifender Operationen in STATA im Mikrozensus. GESIS-ZUMA-Methodenbericht Nr. 2007/03. Mannheim: GESIS-ZUMA.
- Kohler, Ulrich; Kreuter, Frauke (2006). *Datenanalyse mit STATA*. 2. Auflage. München: Oldenburg.
- Lechert, Yvonne; Schimpl-Neimanns, Bernhard (2007). Mikrozensus Scientific Use File 2005 – Dokumentation und Datenaufbereitung. GESIS-ZUMA-Methodenbericht Nr. 2007/08. Mannheim: GESIS-ZUMA, 30.05.08.
URL: http://www.gesis.org/Publikationen/Berichte/ZUMA_Methodenberichte/documents/pdfs/2007/07_08_lechert.pdf
- Lengerer, Andrea; Bohr, Jeanette; Janssen, Andrea (2005). Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen. GESIS-ZUMA-Arbeitsbericht Nr. 2005/05. Mannheim: GESIS-ZUMA, 30.05.08.
URL: http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/05/AB_05_05.pdf
- Lengerer, Andrea; Boehle, Mara (2006). Rekonstruktion von Bandsatzerweiterungen zu Haushalt, Familie und Lebensformen im Mikrozensus. GESIS-ZUMA-Methodenbericht Nr. 2006/05. Mannheim: GESIS-ZUMA.
- Nöthen, Manuela (2005). Von der „traditionellen Familie“ zu „neuen Lebensformen“. Neuerungen in der Familienberichterstattung des Mikrozensus. In: *Wirtschaft und Statistik* 1/2005: S. 25-40.
- Statistisches Bundesamt (Hg., 2003): *Fachserie 1: Bevölkerung und Erwerbstätigkeit, Reihe 3: Haushalte und Familien (Ergebnisse des Mikrozensus)*. 2002. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt (Hg., 2006a): *Pressebroschüre: Leben in Deutschland – Haushalte, Familien und Gesundheit, Ergebnisse des Mikrozensus 2005*. Wiesbaden.

Anhang

Erläuterungen Tabelle 2: Ausprägungen der Variablen der Bevölkerungsbegriffe und Abgrenzungen im Mikrozensus SUF 2005

EF29	=	Erwerbstyp 1 Erwerbstätige
EF30	=	Bevölkerung: Haupt- oder Nebenwohnsitz 1 Bevölkerung am Hauptwohnsitz ohne weiteren Wohnsitz 2 Bevölkerung am Hauptwohnsitz mit weiterem Wohnsitz 3 Bevölkerung am Nebenwohnsitz (mit weiterem Wohnsitz)
EF31	=	Bevölkerung: Privathaushalte 1 Bevölkerung in Privathaushalten
EF504	=	Erwerbstyp 1 Erwerbstätige
EF505	=	Bevölkerung: Haupt- oder Nebenwohnsitz 1 Hauptwohnung ohne Nebenwohnsitz (Bevölkerung am Ort der Hauptwohnung ohne weiteren Wohnsitz) 2 Hauptwohnung mit Nebenwohnsitz (Bevölkerung am Ort der Hauptwohnung mit weiterem Wohnsitz)
EF506	=	Bevölkerung: Privathaushalte 1 Bev. in Privat-HH (Personen, die zur Bevölkerung in Privathaushalten gehören)
EF507	=	Stellung innerhalb des Haushalts 1 Bezugsperson
EF508	=	Bevölkerung: Familienwohnsitz (tradit. Familienkonzept) 1 Bev. am Familienwohnsitz
EF509	=	Stellung innerh. der Familie (tradit. Familienkonzept) 1 Bezugsperson in der Familie
EF513	=	Bevölkerung: Wohnsitz der Lebensgem. (Konzept der Lebensformen) 1 Bev. am Wohns. Lebensgem. (Bevölkerung am Wohnsitz der Lebensgemeinschaft)
EF514	=	Stellung innerh. der Lebensgem. (Konzept der Lebensformen) 1 Bezugsperson in der Lebensgemeinschaft
EF662	=	Haupteinkommensbezieher im Haushalt 1 HEK ist Bezugsperson und selbstständiger Landwirt (Vollzeit) 2 HEK hat allein höchste Einkommensklasse 3 HEK hat mit weiterer Person höchste Einkommensklasse 4 Sonstige Haupteinkommensbezieher(in) (Bezugsperson im Haushalt)
EF761	=	Bevölkerung: Wohnsitz der Familie/Lebensform (Lebensformenkonzept) 1 Bezugsperson der Familie/Lebensform am Hauptwohnsitz, alle weiteren Personen ebenfalls am Hauptwohnsitz 2 Bezugsperson der Familie/Lebensform am Hauptwohnsitz, mindestens eine weitere Person jedoch am Nebenwohnsitz
EF762	=	Stellung innerh. der Familie/Lebensform (Lebensformenkonzept) 2 Ehe- oder Lebenspartner(in) der Bezugsperson in der Familie/Lebensform

- EF763 = Bevölkerung nach Familien oder Lebensformen (Lebensformenkonzept)
3 Nichteheliche oder gleichgeschlechtliche Lebenspartner(in) mit ledigen Kindern im Haushalt
8 Nichteheliche oder gleichgeschlechtliche Lebenspartner(in) ohne ledige Kinder im Haushalt
- EF809 = Familien- oder Lebensformtyp
1 F1 Ehepaar mit ledigen Kindern
2 F2 Nichteheliche Lebensgemeinschaft mit ledigen Kindern
3 F3 Gleichgeschlechtliche Lebensgemeinschaft mit ledigen Kindern
4 F4 Alleinerziehende mit ledigen Kindern
- EF863 = Bevölkerung: Familienwohnsitz (tradit. Familienkonzept)
1 Familien am Hauptwohnsitz (Bezugsperson der Familie am Hauptwohnsitz)
- EF864 = Stellung innerh. der Familie (tradit. Familienkonzept)
1 Bezugsperson in der Familie

Anhang: STATA do-Files

```
version 10
clear
capture log close
*Ins Arbeitsverzeichnis wechseln
**** cd p:\Arbeitsverzeichnis
* Log-Datei
log using bandsatzerweiterung_mz2005.log, text replace
set more off
set mem 500m
set dp comma
```

*1. Replikation der Variable EF664 "Zahl der Erwerbstätigen im Haushalt"

* Einlesen der benötigten Variablen

```
use EF1 EF3 EF4 EF29 EF31 EF661 EF662 EF664 using "Pfadname\mz2005.dta", clear
set more off
```

/*Kommentarblock:

```
* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:
* Bundesland: EF1
* Systemfreie Nr.: Auswahlbezirk: EF3
* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4
```

```
* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"
*           1 "Bevölkerung in Privathaushalten"
*           2 "Bevölkerung in Gemeinschaftsunterkünften"
*           M0 "Entfällt /Haushalt ist leer oder nicht
*              befragt)"
```

```
*           auf die Bevölkerung in Privathaushalten (EF31=1)
```

```
* Einschränkung der Variablen EF29 "Erwerbstyp"
*           1 "Erwerbstätige"
*           2 "Erwerbslose"
*           3 "Arbeitssuchende Nichterwerbspersonen"
*           4 "Sonstige Nichterwerbspersonen"
*           M0 "Entfällt (bei Ausfall, etc.)"
```

```
* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts"
*           1 „Bezugsperson im Haushalt (HBP [erste
*              eingetragene Person])
*           [...]
```

```
*           auf die Bezugsperson im Haushalt (EF661=1)*/
```

*Erster Schritt: Erzeugen der Anzahl der Bezugspersonen im Haushalt

```
egen n_bezper_hh = total(EF31==1 & EF662>=1 & EF662<=4), by(EF1 EF3 EF4)
```

* Label-Vergabe

label variable n_bezper_hh "Anzahl der Bezugspersonen im Haushalt"

tab n_bezper_hh

* Selektion der Haushalte mit nur einer Bezugsperson, Deklaration als Missing

* bei mehr als einer Bezugsperson

```
if n_bezper_hh==1 {  
    egen v664 = sum(EF29==1 & EF31==1), by(EF1 EF3 EF4)  
}  
else {  
    replace v664==.  
}
```

label variable v664 "Replikation: Zahl der Erwerbstätigen im Haushalt"

* Kontrollliste der ersten 25 Fälle

list EF1 EF3 EF4 EF29 EF31 EF664 v664 in 1/25, nolab sepby(EF1 EF3 EF4)

* Überprüfen, ob die Anzahl der Erwerbstätigen mit der Kontrollvariable EF664

* übereinstimmt durch Kontrolle der Diagonalen

tab v664 EF664 if EF31==1, miss

* Zusammenfassen der Zahl der Erwerbstätigen im Haushalt bei sechs Personen und * mehr zu der Kategorie „mindestens 6 Personen“

recode v664 (6/9=6), gen(y664) copyrest

tab EF664 y664, miss

* Generieren der abweichenden Fälle der bei den individuellen Ausprägungen der

* Variable EF664 auf Haushaltsebene

egen f664=sum((y664!=EF664)*EF31==1), by(EF1 EF3 EF4)

* Auflisten der Fälle, in denen die zu rekonstruierende Variable nicht mit der

* Ursprungsvariable übereinstimmt

list EF1 EF3 EF4 EF29 EF661 EF664 y664 if f664>0 & EF664<99 & EF31==1, nolab sepby(EF1 EF3 EF4)

* Auszählen der abweichenden Fälle

count if f664>0 & EF664<99 & EF31==1

* 17 Fälle sind im Mikrozensus 2005 inkonsistent

* Auszählen der durch abweichende Daten betroffenen Haushalte

count if f664>0 & EF664<99 & EF31==1 & EF661==1

* 6 Haushalte sind inkonsistent, aber in 2 Haushalten liegt sogar keine

* Bezugsperson vor, d.h. insgesamt sind 8 Haushalte inkonsistent

*2. Replikation der Variable EF713 "Zahl der nichtehelichen Lebensgemeinschaften im Haushalt"

* Einlesen der benötigten Variablen

use EF1 EF3 EF4 EF27 EF31 EF35 EF661 EF713 EF761 EF809 using "Pfadname\mz2005.dta", clear
set more off

/*Kommentarblock:

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:

* Bundesland: EF1

* Systemfreie Nr.: Auswahlbezirk: EF3

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"

* 1 "Bevölkerung in Privathaushalten"

* 2 "Bevölkerung in Gemeinschaftsunterkünften"

* M0 "Entfällt /Haushalt ist leer oder nicht

befragt)"

* auf die Bevölkerung in Privathaushalten (EF31=1)

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform
(Lebensformenkonzept)"

* 1 "Bezugsperson in der Familie/Lebensform"

* 2 "Ehefrau oder Lebenspartner(in) der
Bezugsperson in der Familie/Lebensform"

* 3 "Lediges Kind in der Familie"

* 4 "Sonstige Nichterwerbspersonen"

* M0 "Entfällt (Bevölkerung in
Gemeinschaftsunterkünften)"

* auf die Bezugsperson in der Familie/Lebensform
(EF35=1)

* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts“

* 1 „Bezugsperson im Haushalt (HBP [erste
eingetragene Person])

* [...]

* auf die Bezugsperson im Haushalt (EF661=1)

* Einschränkung der Variablen EF809 "Familien- oder Lebensformtyp"

* 1 "F1 Ehepaar mit ledigen Kindern"

* 2 "F2 Nichteheliche Lebensgemeinschaft mit
ledigen Kindern"

* 3 "F3 Gleichgeschlechtliche Lebensgemeinschaft * mit ledigen Kindern"

* 4 "F4 Alleinerziehende mit ledigen Kindern"

* 5 "F5 Ehepaar ohne ledige Kinder"

* 6 "F6 Nichteheliche Lebensgemeinschaft ohne
ledige Kinder"

* 7 "F7 Gleichgeschlechtliche Lebensgemeinschaft * ohne ledige Kinder"

* 8 "F8 Alleinstehende"

* M0 "Entfällt (Bevölkerung in
Gemeinschaftsunterkünften, bekannte
Ausfälle, Fortzüge, Gestorbene,
unbekannte Abgänge, leer stehende
Wohnungen)

* auf Nichteheliche Lebensgemeinschaften (EF809=2 | EF809=6) */

egen n_bezper_hh_lebensform = total(EF31==1 & EF35==1), by(EF1 EF3 EF4 EF27)

label variable n_bezper_hh_lebensform "Anzahl der Bezugspersonen im Haushalt"

tab n_bezper_hh_lebensform

*Betrachten der Lebensformen mit nur einer Bezugsperson

```
if n_bezper_hh_lebensform==1 {  
    egen v713 = total(EF31==1 & EF35==1 & (EF809==2 | EF809==6)), by(EF1 EF3 EF4 EF27)  
}  
else {  
    replace v713==.  
}  
}
```

label variable v713 "Zahl der nichtehelichen Lebensgemeinschaften im Haushalt"

* Kontrolle der Diagonalen durch Kreuztabellierung

tab v713 EF713 if EF31==1, miss

* Kontrollliste der ersten 25 Fälle getrennt nach Haushalten

list EF1 EF3 EF4 EF27 EF31 EF35 EF809 EF713 v713 in 1/25, nolab sepby(EF1 EF3 EF4)

*Nicht passende Fälle untersuchen

list EF1 EF3 EF4 EF27 EF31 EF35 EF809 EF713 v713 if v713==1 & v713!=EF713,nolab sepby(EF1 EF3 EF4 EF27)

* Bilden der aggregierten Summe der abweichenden Fälle hinsichtlich der

* Ausprägung der Variablen EF713 auf Haushaltsebene

```
gen y713=v713  
egen f713=sum((y713!=EF713)*EF31==1), by(EF1 EF3 EF4)
```

* Kontrollliste

list EF1 EF3 EF4 EF31 EF35 EF809 EF661 EF713 y713 if f713>0 & EF713<99 & EF31==1, nolab sepby(EF1 EF3 EF4)

* Auszählen der abweichenden Fälle

count if f713>0 & EF713<99 & EF31==1

* 1275 Fälle sind problematisch

* Auszählen der abweichenden Haushalte

count if f713>0 & EF713<99 & EF31==1 & EF661==1

* 314 Haushalte sind problematisch

*3. Replikation der Variable EF736 "Erwerbstyp: Haupteinkommensbezieher"

* Einlesen der benötigten Daten

use EF1 EF3 EF4 EF31 EF35 EF661 EF662 EF736 using "Pfadname\mz2005.dta", clear
set more off

/*Kommentarblock:

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:

* Bundesland: EF1

* Systemfreie Nr.: Auswahlbezirk: EF3

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4

* EF29 "Erwerbstyp"

* 1 "Erwerbstätige"

* 2 "Erwerbslose"

* 3 "Arbeitssuchende Nichterwerbspersonen"

* 4 "Sonstige Nichterwerbspersonen"

* M0 "Entfällt (bei Ausfall)"

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"

* 1 "Bevölkerung in Privathaushalten"

* 2 "Bevölkerung in Gemeinschaftsunterkünften"

* M0 "Entfällt /Haushalt ist leer oder nicht befragt)"

* auf die Bevölkerung in Privathaushalten (EF31=1)

* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts"

* 1 „Bezugsperson im Haushalt (HBP [erste eingetragene Person])
[...]

* auf die Bezugsperson im Haushalt (EF661=1)

* Einschränkung der Variablen EF662 "Haupteinkommensbezieher im Haushalt"

* 1 " HEK ist Bezugsperson und selbstständiger Landwirt (Vollzeit)"

* 2 " HEK hat allein höchste Einkommensklasse"

* 3 " HEK hat mit weiterer Person höchste Einkommensklasse"

* 4 " Sonstige Haupteinkommensbezieher(in) (Bezugsperson im Haushalt)"

* 5 " Person im Haushalt mit Haupteinkommensbezieher(in) als selbstständige(r) Landwirt(in) (Vollzeit)"

* 6 " Person mit höchster Einkommensklasse, aber * nicht HEK"

* 7 " Person mit Einkommensangabe, aber nicht höchste Einkommensklasse"

* 8 " Person ohne Einkommensangabe, aber andere * Haushaltsmitglieder

mit

* Angabe des Individualeinkommens"

* 9 " Person ohne Einkommensangabe, alle anderen * Haushaltsmitglieder

ebenfalls

* ohne Angabe des Individualeinkommens und Person nicht Bezugsperson des Haushalts"

* M 0 " Entfällt (Bevölkerung in Gemeinschaftsunterkünften)"

* auf EF662=1-4 */

- * Generieren der abgeleiteten Variable v736 für Haushalte mit nur einer Haushaltsbezugsperson

```
if n_bezper_hh==1 {  
  egen v736 = max( EF29* (EF31==1 & (EF662==1 | EF662==2 | EF662==3 | EF662==4))), by(EF1 EF3 EF4)  
}  
else {  
  replace v736==.  
}
```

- * Kreuztabellierung der abgeleiteten Variable v736 mit der Ursprungsvariable EF736 und Kontrolle der korrekten Zuweisung anhand der Diagonalen

```
tab v736 EF736 if EF31==1, miss
```

- * Kontrollliste der ersten 25 Fälle

```
list EF1 EF3 EF4 EF29 EF31 EF662 EF736 v736 in 1/25, nolab sepby(EF1 EF3 EF4)
```

- * Bilden der aggregierten Summe der abweichenden Fälle hinsichtlich der Ausprägung der Variablen EF736 auf Haushaltsebene

```
gen y736=v736  
egen f736=sum((y736!=EF736)*EF31==1), by(EF1 EF3 EF4)
```

- * Kontrollliste mitsamt der aggregierten abweichenden Fälle

```
list EF1 EF3 EF4 EF31 EF661 EF662 EF736 y736 if f736>0 & EF736<99 & EF31==1, nolab sepby(EF1 EF3 EF4)
```

- * Auszählen der abweichenden Fälle

```
count if f736>0 & EF736<99 & EF31==1
```

- * 108 Fälle sind problematisch

- * Auszählen der abweichenden Haushalte

```
count if f736>0 & EF736<99 & EF31==1 & (EF662==1 | EF662==2 | EF662==3 | EF662==4)
```

- * 50 Haushalte sind problematisch

*4. Replikation der Variable EF758 "Erwerbstyp: Haushaltsbezugsperson"

* Einlesen der benötigten Daten

use EF1 EF3 EF4 EF29 EF31 EF661 EF758 using "Pfadname\mz2005.dta", clear
set more off

/* Kommentarblock:

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:

* Bundesland: EF1

* Systemfreie Nr.: Auswahlbezirk: EF3

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4

* EF29 "Erwerbstyp"

* 1 "Erwerbstätige"

* 2 "Erwerbslose"

* 3 "Arbeitssuchende Nichterwerbspersonen"

* 4 "Sonstige Nichterwerbspersonen"

* M0 "Entfällt (bei Ausfall)"

*

*

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"

* 1 "Bevölkerung in Privathaushalten"

* 2 "Bevölkerung in Gemeinschaftsunterkünften"

* M0 "Entfällt /Haushalt ist leer oder nicht

befragt)"

*

* auf die Bevölkerung in Privathaushalten (EF31=1)

*

* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts"

* 1 „Bezugsperson im Haushalt (HBP [erste

eingetragene Person])

* [...]

*

* auf die Bezugsperson im Haushalt (EF661=1)*/

* Rekonstruktion der abgeleiteten Variablen bei Haushalten mit nur einer

* Haushaltsbezugsperson

```
if n_bezper_hh==1 {  
    egen v758 = max( EF29* (EF31==1 & EF661==1)), by(EF1 EF3 EF4)  
}  
else {  
    replace v758==.  
}  
}
```

* Kontrolle der Diagonalen durch Kreuztabellierung

tab v758 EF758

* Kontrollliste der ersten 25 Fälle

list EF1 EF3 EF4 EF29 EF31 EF661 EF758 v758 in 1/25, nolab sepby(EF1 EF3 EF4)

* Bilden der aggregierten Summe der abweichenden Fälle hinsichtlich der

* Ausprägung der Variablen EF758 auf Haushaltsebene

```
gen y758=v758
```

```
egen f758=sum((y758!=EF758)*EF31==1), by(EF1 EF3 EF4)
```

* Kontrollliste mitsamt der aggregierten abweichenden Fälle

```
list EF1 EF3 EF4 EF31 EF661 EF758 y758 if f758>0 & EF758<99 & EF31==1, nolab sepby(EF1 EF3 EF4)
```

* Auszählen der abweichenden Fälle

```
count if f758>0 & EF758<99 & EF31==1
```

* 137 Fälle sind problematisch

* Auszählen der abweichenden Haushalte

```
count if f758>0 & EF758<99 & EF31==1 & EF661==1
```

* 60 Haushalte sind problematisch

***5. Replikation der Variable EF770 "Zahl der ledigen Kinder in der Familie/Lebensform insgesamt (ohne Altersbegrenzung)"**

* Einlesen der benötigten Daten

use EF1 EF3 EF4 EF27 EF31 EF35 EF44 EF661 EF770 using "Pfadname\mz2005.dta", clear
set more off

/* Kommentarblock:

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:

* Bundesland: EF1

* Systemfreie Nr.: Auswahlbezirk: EF3

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"

* 1 "Bevölkerung in Privathaushalten"
* 2 "Bevölkerung in Gemeinschaftsunterkünften"
* M0 "Entfällt /Haushalt ist leer oder nicht
* befragt)"

* auf die Bevölkerung in Privathaushalten (EF31=1)

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform"

* 1 "Bezugsperson in der Familie/Lebensform"
* 2 "Ehefrau oder Lebenspartner(in) der
* Bezugsperson in der Familie/Lebensform"
* 3 "Lediges Kind in der Familie"
* M0 "Entfällt (Bevölkerung in
* Gemeinschaftsunterkünften)"

* Ohne Altersbegrenzung EF44 "Alter"

* 0 "0 Jahre "
* 95 " (95 Jahre und älter)"

* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts“

* 1 „Bezugsperson im Haushalt (HBP [erste
* eingetragene Person])
* [...]

* auf die Bezugsperson im Haushalt (EF661=1)*/

* Vergleiche Beispiel 2 (oder auch Kapitel 3.2): Zuerst muss die Anzahl der

* Bezugspersonen in der Lebensform gebildet werden

egen n_bezper_hh_lebensform = total(EF31==1 & EF35==1), by(EF1 EF3 EF4 EF27)

* Rekonstruktion der abgeleiteten Variablen für Lebensformen mit nur einer

* Bezugsperson

```
if n_bezper_hh_lebensform==1 {  
    egen v770 = total(EF31==1 & EF35==3), by(EF1 EF3 EF4 EF27)  
}  
else {  
    replace v770==.  
}
```

* Kontrolle der Diagonalen durch Kreuztabellierung

tab v770 EF770 if EF31==1, mis

* Kontrollliste der ersten 25 Fälle

list EF1 EF3 EF4 EF31 EF35 EF770 v770 in 1/25, nolab sepby(EF1 EF3 EF4 EF27)

* Bilden der aggregierten Summe der abweichenden Fälle hinsichtlich der

* Ausprägung der Variablen EF770 auf Familie-/ Lebensformebene

gen y770=v770

egen f770=sum((y770!=EF770)*EF31==1), by(EF1 EF3 EF4 EF27)

* Kontrollliste mitsamt der aggregierten abweichenden Fälle

list EF1 EF3 EF4 EF27 EF31 EF35 EF661 EF770 y770 if f770>0 & EF770<99 & EF31==1, nolab sepby(EF1 EF3 EF4 EF27)

* Auszählen der abweichenden Fälle

count if f770>0 & EF770<99 & EF31==1

* 117 Fälle sind problematisch

* Auszählen der abweichenden Familien bzw. Lebensformen

count if f770>0 & EF770<99 & EF31==1 & EF35==3

* 96 Familien/Lebensformen sind problematisch

*6. Replikation der Variable EF824 "Erwerbstyp der Bezugsperson in der Familie/Lebensform"

* Einlesen der benötigten Variablen

use EF1 EF3 EF4 EF27 EF31 EF35 EF661 EF824 using "Pfadname\mz2005.dta", clear

set more off

/* Kommentarblock:

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:

* Bundesland: EF1

* Systemfreie Nr.: Auswahlbezirk: EF3

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27

```
*
*           EF29 "Erwerbstyp"
*           1 "Erwerbstätige"
*           2 "Erwerbslose"
*           3 "Arbeitssuchende Nichterwerbspersonen"
*           4 "Sonstige Nichterwerbspersonen"
*           M0 "Entfällt (bei Ausfall)"
*
```

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"

```
*           1 "Bevölkerung in Privathaushalten"
*           2 "Bevölkerung in Gemeinschaftsunterkünften"
*           M0 "Entfällt /Haushalt ist leer oder nicht
*             befragt)"
*
```

* auf die Bevölkerung in Privathaushalten (EF31=1)

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform"

```
*           1 "Bezugsperson in der Familie/Lebensform"
*           2 "Ehefrau oder Lebenspartner(in) der
*             Bezugsperson in der Familie/Lebensform"
*           3 "Lediges Kind in der Familie"
*           M0 "Entfällt (Bevölkerung in
*             Gemeinschaftsunterkünften)"
*
```

* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts“

```
*           1 „Bezugsperson im Haushalt (HBP [erste
*             eingetragene Person])
*           [...]
*
```

* auf die Bezugsperson im Haushalt (EF661=1)*/

* Erzeugen der Anzahl der Bezugspersonen in der Lebensform

egen n_bezper_hh_lebensform = total(EF31==1 & EF35==1), by(EF1 EF3 EF4 EF27)

* Rekonstruktion der abgeleiteten Variablen für Lebensformen mit nur einer

* Bezugsperson

```
if n_bezper_hh_lebensform==1 {
    egen v824 = max( EF29* (EF31==1 & EF35==1)), by(EF1 EF3 EF4 EF27)
}
else {
    replace v824==.
}
```


}

* Kontrolle der Diagonalen durch Kreuztabellierung

tab v824 EF824

* Kontrollliste der ersten 25 Fälle

list EF1 EF3 EF4 EF29 EF31 EF35 EF824 v824 in 1/25, nolab sepby(EF1 EF3 EF4 EF27)

* Bilden der aggregierten Summe der abweichenden Fälle hinsichtlich der

* Ausprägung der Variablen EF824 auf Familie-/ Lebensformebene

gen y824=v824

egen f824=sum((y824!=EF824)*EF31==1), by(EF1 EF3 EF4 EF27)

* Kontrollliste mitsamt der aggregierten abweichenden Fälle

list EF1 EF3 EF4 EF27 EF29 EF31 EF35 EF824 y824 if f824>0 & EF824<99 & EF31==1, nolab sepby(EF1 EF3 EF4 EF27)

* Auszählen der abweichenden Fälle

count if f824>0 & EF824<99 & EF31==1

* 144 Fälle sind falsch

* Auszählen der abweichenden Haushalte

count if f824>0 & EF824<99 & EF31==1 & EF661==1

* 62 Haushalte sind falsch

* Auszählen der abweichenden Familien bzw. Lebensformen

count if f824>0 & EF824<99 & EF31==1 & EF35==3

* 47 abweichende Familien bzw. Lebensformen sind identifizierbar

*7. Replikation der Variable EF848 "Erwerbstyp des Ehe- oder Lebenspartners"

* Einlesen der benötigten Variablen

```
use EF1 EF3 EF4 EF27 EF29 EF31 EF35 EF661 EF848 using "Pfadname\mz2005.dta", clear
```

```
set more off
```

/* Kommentarblock:

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:

* Bundesland: EF1

* Systemfreie Nr.: Auswahlbezirk: EF3

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27

```
*                               EF29 "Erwerbstyp"
```

```
*                               1 "Erwerbstätige"
```

```
*                               2 "Erwerbslose"
```

```
*                               3 "Arbeitssuchende Nichterwerbspersonen"
```

```
*                               4 "Sonstige Nichterwerbspersonen"
```

```
*                               M0 "Entfällt (bei Ausfall)"
```

```
* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"
```

```
*                               1 "Bevölkerung in Privathaushalten"
```

```
*                               2 "Bevölkerung in Gemeinschaftsunterkünften"
```

```
*                               M0 "Entfällt /Haushalt ist leer oder nicht  
                               befragt)"
```

```
*                               auf die Bevölkerung in Privathaushalten (EF31=1)
```

```
* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform"
```

```
*                               1 "Bezugsperson in der Familie/Lebensform"
```

```
*                               2 "Ehefrau oder Lebenspartner(in) der  
                               Bezugsperson in der Familie/Lebensform"
```

```
*                               3 "Lediges Kind in der Familie"
```

```
*                               M0 "Entfällt (Bevölkerung in  
                               Gemeinschaftsunterkünften)"
```

```
* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts"
```

```
*                               1 „Bezugsperson im Haushalt (HBP [erste  
                               eingetragene Person])
```

```
*                               [...]
```

```
*                               auf die Bezugsperson im Haushalt (EF661=1)*/
```

* Erzeugen der Anzahl der Bezugspersonen in der Lebensform

```
egen n_bezper_hh_lebensform = total(EF31==1 & EF35==1), by(EF1 EF3 EF4 EF27)
```

* Replikation der abgeleiteten Variablen

```
if n_bezper_hh_lebensform==1 {  
    egen v848 = max( EF29* (EF31==1 & EF35==2)), by(EF1 EF3 EF4 EF27)
```

```
}
```

```
else {  
    replace v848==.
```

```
}
```

* Kontrolle der Diagonalen anhand der Kreuztabellierung

tab v848 EF848

* Kontrollliste

list EF1 EF3 EF4 EF27 EF29 EF31 EF35 EF848 v848 in 1/25, nolab sepby(EF1 EF3 EF4 EF27)

* Bilden der aggregierten Summe der abweichenden Fälle hinsichtlich der

* Ausprägung der Variablen EF848 auf Familie-/ Lebensformebene

gen y848=v848

egen f848=sum((y848!=EF848)*EF31==1), by(EF1 EF3 EF4 EF27)

* Kontrollliste mitsamt der aggregierten abweichenden Fälle

list EF1 EF3 EF4 EF27 EF29 EF31 EF35 EF848 y848 if f848>0 & EF31==1, nolab sepby(EF1 EF3 EF4 EF27)

* Auszählen der abweichenden Fälle

count if f848>0 & EF848<99 & EF31==1

* 67 Fälle weichen ab

* Auszählen der abweichenden Haushalte

count if f848>0 & EF848<99 & EF31==1 & EF661==1

* 28 Haushalte weichen ab

* Auszählen der abweichenden Familien bzw. Lebensformen

count if f848>0 & EF848<99 & EF31==1 & EF35==3

* 13 Familien bzw. Lebensformen weichen ab

*8. Replikation der Variable EF866 "Zahl der ledigen Kinder insgesamt in der Familie"

* Einlesen der benötigten Variablen

```
use EF1 EF3 EF4 EF25 EF31 EF34 EF35 EF661 EF866 using "Pfadname\mz2005.dta", clear
set more off
```

/* Kommentarblock:

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:

* Bundesland: EF1

* Systemfreie Nr.: Auswahlbezirk: EF3

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4

* Laufende Nr.: Familie im Haushalt (trad. Familienkonzept): EF25

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte"

```
*      1 "Bevölkerung in Privathaushalten"
*      2 "Bevölkerung in Gemeinschaftsunterkünften"
*      M0 "Entfällt /Haushalt ist leer oder nicht
*          befragt)"
*
```

```
*      auf die Bevölkerung in Privathaushalten (EF31=1)
*
```

```
*      EF34 "Stellung innerh. der Familie (tradit.
*          Familienkonzept)"
*
```

```
*      1 "Bezugsperson in der Familie"
*      2 "Ehefrau der Bezugsperson in der Familie"
*      3 "Lediges Kind in der Familie"
*      M 0 "Entfällt (Bevölkerung in
*          Gemeinschaftsunterkünften)"
*
```

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform"

```
*      1 "Bezugsperson in der Familie/Lebensform"
*      2 "Ehefrau oder Lebenspartner(in) der
*          Bezugsperson in der Familie/Lebensform"
*      3 "Lediges Kind in der Familie"
*      M0 "Entfällt (Bevölkerung in
*          Gemeinschaftsunterkünften)"
*
```

* Einschränkung der Variablen EF661 „Stellung innerhalb des Haushalts“

```
*      1 „Bezugsperson im Haushalt (HBP [erste
*          eingetragene Person])
*      [...]
*
```

```
*      auf die Bezugsperson im Haushalt (EF661=1)*/
```

* Erzeugen der Anzahl der Bezugspersonen in der Familie

```
egen n_bezper_hh_trad = total(EF31==1 & EF35==1), by(EF1 EF3 EF4 EF25)
```

label variable n_bezper_hh_trad "Anzahl der Bezugspersonen im Haushalt"

tab n_bezper_hh_trad

* Replikation der abgeleiteten Variablen

```
if n_bezper_hh_trad==1 {
    egen v866 = total(EF31==1 & EF34==3 | EF35==3), by(EF1 EF3 EF4 EF25)
}
```

```
else {  
  replace v866==.  
}
```

* Kontrolle der Diagonalen anhand der Kreuztabellierung

```
tab v866 EF866
```

* Kontrollliste

```
list EF1 EF3 EF4 EF25 EF31 EF34 EF35 EF661 EF866 v866 in 1/25, nolab sepby(EF1 EF3 EF4 EF25)
```

* Bilden der aggregierten Summe der abweichenden Fälle hinsichtlich der

* Ausprägung der Variablen EF866 auf Familieebene

```
gen y866=v866  
egen f866=sum((y866!=EF866)*EF31==1), by(EF1 EF3 EF4 EF25)
```

* Kontrollliste mitsamt der aggregierten abweichenden Fälle

```
list EF1 EF3 EF4 EF25 EF31 EF34 EF35 EF661 EF866 y866 if f866>0 & EF31==1, nolab sepby(EF1 EF3 EF4 EF25)
```

* Auszählen der abweichenden Fälle

```
count if f866>0 & EF866<99 & EF31==1
```

* 119 Fälle weichen ab

* Auszählen der abweichenden Haushalte

```
count if f866>0 & EF866<99 & EF31==1 & EF661==1
```

* 12 Haushalte weichen ab

* Auszählen der abweichenden Familien

```
count if f866>0 & EF866<99 & EF31==1 & EF34==3
```

* 97 Familien weichen ab

Anhang: SPSS sps-Files

* Einlesen der benötigten Variablen

```
GET FILE='mz2005.sav'  
/keep EF1 EF3 EF4 EF1 EF3 EF4 EF25 EF27 EF29 EF30 EF31 EF38 EF661 EF662 EF664 EF713 EF736 EF758 EF770 EF824  
EF848 EF866.
```

*1. Replikation der Variable EF664 "Zahl der Erwerbstätigen im Haushalt".

/* Kommentarblock.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".

* 2 "Bevölkerung in Gemeinschaftsunterkünften".

* M0 "Entfällt /Haushalt ist leer oder nicht.

* befragt)".

*
* auf die Bevölkerung in Privathaushalten (EF31=1).

* Einschränkung der Variablen EF29 "Erwerbstyp".

* 1 "Erwerbstätige".

* 2 "Erwerbslose".

* 3 "Arbeitssuchende Nichterwerbspersonen".

* 4 "Sonstige Nichterwerbspersonen".

* M0 "Entfällt (bei Ausfall, etc.)" . */

sort cases by EF1 EF3 EF4.

compute x664=0.

if (EF29=1 & EF31=1) x664=1.

aggregate outfile=*

 mode=addvariables

 /presorted

 /break EF1 EF3 EF4

 /v664'Replikation: Zahl der Erwerbstätigen im Haushalt'= sum(x664).

crosstabs ef664 by v664

/missing include.

*2. Replikation der Variable EF713 "Zahl der nichtehelichen Lebensgemeinschaften im Haushalt".

/* Kommentarblock.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27.

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".
 * 2 "Bevölkerung in Gemeinschaftsunterkünften".
 * M0 "Entfällt /Haushalt ist leer oder nicht.
 * befragt)".

* auf die Bevölkerung in Privathaushalten (EF31=1).

* Einschränkung der Variablen EF35 "Stellung.

* innerhalb der Familie/Lebensform.

* (Lebensformenkonzept)".

* 1 "Bezugsperson in der Familie/Lebensform".

* 2 "Ehefrau oder Lebenspartner(in) der.
 * Bezugsperson in der Familie/Lebensform".

* 3 "Lediges Kind in der Familie".

* 4 "Sonstige Nichterwerbspersonen".

* M0 "Entfällt (Bevölkerung in.

* Gemeinschaftsunterkünften)".

* auf die Bezugsperson in der Familie/Lebensform

* (EF35=1).

* Einschränkung der Variablen EF809 "Familien- oder Lebensformtyp".

* 1 "F1 Ehepaar mit ledigen Kindern".

* 2 "F2 Nichteheliche Lebensgemeinschaft mit.
 * ledigen Kindern".

* 3 "F3 Gleichgeschlechtliche.

* Lebensgemeinschaft.

* mit ledigen Kindern".

* 4 "F4 Alleinerziehende mit ledigen Kindern".

* 5 "F5 Ehepaar ohne ledige Kinder".

* 6 "F6 Nichteheliche Lebensgemeinschaft ohne.
 * ledige Kinder".

* 7 "F7 Gleichgeschlechtliche.

* Lebensgemeinschaft.

* ohne ledige Kinder".

* 8 "F8 Alleinstehende".

* M0 "Entfällt (Bevölkerung in.

* Gemeinschaftsunterkünften, bekannte.

* Ausfälle, Fortzüge, Gestorbene.

* unbekannte Abgänge, leer stehende.

* Wohnungen).

* auf Nichteheliche Lebensgemeinschaften (EF809=2 |. *

EF809=6).*/

sort cases by EF1 EF3 EF4.

compute x713=0.

if (EF31=1 & EF35=1 & EF762=1 & (EF761>=1 | EF761<=2) & (EF809=2 | EF809=6)) x713=1.

aggregate outfile=*

mode=addvariables

/presorted

/break EF1 EF3 EF4

/v713'Replikation: Zahl der nichtehelichen Lebensgemeinschaften'= sum(x713).

RECODE

v713 (2=1).

EXECUTE.

crosstabs ef713 by v713

/missing include.

*3. Replikation der Variable EF736 "Erwerbstyp: Haupteinkommensbezieher".

/* Kommentarblock:.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27.

* EF29 "Erwerbstyp".

* 1 "Erwerbstätige".

* 2 "Erwerbslose".

* 3 "Arbeitssuchende Nichterwerbspersonen".

* 4 "Sonstige Nichterwerbspersonen".

* M0 "Entfällt (bei Ausfall)".

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".

* 2 "Bevölkerung in Gemeinschaftsunterkünften".

* M0 "Entfällt /Haushalt ist leer oder nicht.

* befragt)".

* auf die Bevölkerung in Privathaushalten (EF31=1).

* Einschränkung der Variablen EF662 "Haupteinkommensbezieher im Haushalt".

* 1 "HEK ist Bezugsperson und selbstständiger.

* Landwirt (Vollzeit)".

* 2 "HEK hat allein höchste Einkommensklasse".

* 3 "HEK hat mit weiterer Person höchste.

* Einkommensklasse".

* 4 "Sonstige Haupteinkommensbezieher(in).

* (Bezugsperson im Haushalt)".

* 5 "Person im Haushalt mit.

* Haupteinkommensbezieher(in) als.

* selbstständige(r).

* Landwirt(in) (Vollzeit)".

* 6 "Person mit höchster Einkommensklasse.

* aber nicht HEK".

* 7 "Person mit Einkommensangabe, aber nicht.

* höchste Einkommensklasse".

* 8 "Person ohne Einkommensangabe, aber andere. *

* Angabe des Individualeinkommens".

* 9 "Person ohne Einkommensangabe, alle anderen. *

Haushaltsmitglieder mit.

ebenfalls.

* ohne Angabe des Individualeinkommens und.

* Person nicht Bezugsperson des.

* Haushalts".

* M 0 " Entfällt (Bevölkerung in.

* Gemeinschaftsunterkünften)".

* auf EF662=1-4.*/

* Eingrenzung der Bevölkerung auf Personen in Privathaushalten.

SELECT IF EF31=1.

* Fälle sortieren nach der Haushaltsnummer.

SORT CASES BY EF1 EF3 EF4 EF27.

compute HEK=ef662.

RECODE

HEK (1 thru 4=1) (5 thru 9=0).

EXECUTE.

freq HEK.

* Erwerbstyp (EF29) für jeden Haupteinkommensbezieher (HEK=1) in file schreiben.

TEMPORARY.

SELECT IF HEK=1.

SAVE OUTFILE='<Pfadname:>\V736.sav'

/ KEEP=EF1 EF3 EF4 EF27 EF29

/ RENAME EF29=V736.

* Jeder Person eines Haushalts im Ausgangsfile (*) den Erwerbstyp (EF29) des jeweiligen Haupteinkommensbeziehers (HEK) zuspielen.

MATCH FILES TABLE='<Pfadname:>\V736.sav'

/ FILE=*

/ BY EF1 EF3 EF4 EF27.

* Kontrolle.

CROSSTABS/TABLES=V736 BY EF736.

* Jeder Person eines Haushalts im Ausgangsfile (*) den Erwerbstyp (EF29) des jeweiligen Haupteinkommensbeziehers zuspielen.

MATCH FILES TABLE='<Pfadname\V736.sav'

/ FILE=*

/ BY EF3 EF4.

* Kontrolle.

CROSSTABS/TABLES=V736 BY EF736.

***4. Replikation der Variable EF758 "Erwerbstyp: Haushaltsbezugsperson".**

/* Kommentarblock:.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* EF29 "Erwerbstyp".

* 1 "Erwerbstätige".

* 2 "Erwerbslose".

* 3 "Arbeitssuchende Nichterwerbspersonen".

* 4 "Sonstige Nichterwerbspersonen".

* M0 "Entfällt (bei Ausfall)".

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".

* 2 "Bevölkerung in Gemeinschaftsunterkünften".

* M0 "Entfällt /Haushalt ist leer oder nicht

befragt)".

* auf die Bevölkerung in Privathaushalten (EF31=1).

* Einschränkung der Variablen EF661 "Familien- oder Lebensformtyp".

* 1 " 1 Bezugsperson im Haushalt (HBP [erste eingetragene Person])".

* M 0 " Entfällt (Bevölkerung in

Gemeinschaftsunterkünften)". */

sort cases by EF1 EF3 EF4.

compute x758=0.

if (EF31=1 & EF661=1) x758=EF29.

aggregate outfile=*mode=addvariables

/presorted

/break EF1 EF3 EF4

/v758 'Replikation: Erwerbstyp der Haushaltsbezugsperson' = sum(x758).

crosstabs ef758 by v758

/missing include.

***5. Replikation der Variable EF770 "Zahl der ledigen Kinder in der Familie/Lebensform insgesamt (ohne Altersbegrenzung)".**

/* Kommentarblock:.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27.

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".

* 2 "Bevölkerung in Gemeinschaftsunterkünften".

* M0 "Entfällt /Haushalt ist leer oder nicht.

* befragt)".

*

* auf die Bevölkerung in Privathaushalten (EF31=1).

*

*

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform".

* 1 "Bezugsperson in der Familie/Lebensform".

* 2 "Ehefrau oder Lebenspartner(in) der
 Bezugsperson in der Familie/Lebensform".

* 3 "Lediges Kind in der Familie.

* M0 "Entfällt (Bevölkerung in
 Gemeinschaftsunterkünften)".

*

*

*

* Ohne Altersbegrenzung EF44 "Alter".

* 0 " 0 Jahre ".

* 95 " (95 Jahre und älter)". */

sort cases by EF1 EF3 EF4 EF27.

select if EF31=1.

compute x770=0.

if EF35=3 x770=1.

aggregate outfile=*
mode=addvariables

/presorted

/break EF1 EF3 EF4 EF27

/v770'Replikation: Zahl der ledigen Kinder in der Familie/Lebensform'= sum(x770).

crosstabs ef770 by v770

/missing include.

***6. Replikation der Variable EF824 "Erwerbstyp der Bezugsperson in der Familie/Lebensform".**

/* Kommentarblock:.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27.

* EF29 "Erwerbstyp".

* 1 "Erwerbstätige".

* 2 "Erwerbslose".

* 3 "Arbeitssuchende Nichterwerbspersonen".

* 4 "Sonstige Nichterwerbspersonen".

* M0 "Entfällt (bei Ausfall)".

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".

* 2 "Bevölkerung in Gemeinschaftsunterkünften".

* M0 "Entfällt /Haushalt ist leer oder nicht

befragt)".

* auf die Bevölkerung in Privathaushalten (EF31=1).

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform".

* 1 "Bezugsperson in der Familie/Lebensform".

* 2 "Ehefrau oder Lebenspartner(in) der
Bezugsperson in der Familie/Lebensform".

* 3 "Lediges Kind in der Familie.

* M0 "Entfällt (Bevölkerung in

Gemeinschaftsunterkünften)". */

sort cases by EF1 EF3 EF4 EF27.

select if EF31=1.

compute x824=0.

if (EF31=1 & EF35=1) x824=EF29.

sort cases by EF1 EF3 EF4 EF27.

aggregate outfile=*

mode=addvariables

/presorted

/break EF1 EF3 EF4 EF27

/v824'Replikation: Erwerbstyp der Bezugsperson in der Familie/Lebensform'= sum(x824).

crosstabs ef824 by v824

/missing include.

***7. Replikation der Variable EF848 "Erwerbstyp des Ehe- oder Lebenspartners".**

/* Kommentarblock:.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* Laufende Nr.: Familie/Lebensform im Haushalt (Lebensformenkonzept): EF27.

* EF29 "Erwerbstyp".

* 1 "Erwerbstätige".

* 2 "Erwerbslose".

* 3 "Arbeitssuchende Nichterwerbspersonen".

* 4 "Sonstige Nichterwerbspersonen".

* M0 "Entfällt (bei Ausfall)".

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".

* 2 "Bevölkerung in Gemeinschaftsunterkünften".

* M0 "Entfällt /Haushalt ist leer oder nicht.

* befragt)".

* auf die Bevölkerung in Privathaushalten (EF31=1).

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform".

* 1 "Bezugsperson in der Familie/Lebensform"

* 2 "Ehefrau oder Lebenspartner(in) der

* Bezugsperson in der Familie/Lebensform".

* 3 "Lediges Kind in der Familie.

* M0 "Entfällt (Bevölkerung in.

* Gemeinschaftsunterkünften)".*/

sort cases by EF1 EF3 EF4 EF27.

select if EF31=1.

compute x848=0.

if (EF31=1 & EF35=2) x848=EF29.

sort cases by EF1 EF3 EF4 EF27.

aggregate outfile=*

mode=addvariables

/presorted

/break EF1 EF3 EF4 EF27

/v848'Replikation: Erwerbstyp der Bezugsperson in der Familie/Lebensform'= sum(x848).

crosstabs ef848 by v848

/missing include.

***8. Replikation der Variable EF866 "Zahl der ledigen Kinder insgesamt in der Familie".**

/* Kommentarblock:.

* Arbeitsfile (MZ-2005-Variablenauswahl) sortieren nach Ordnungsnummern:.

* Bundesland: EF1.

* Systemfreie Nr.: Auswahlbezirk: EF3.

* Systemfreie Nr.: Haushalt im Auswahlbezirk: EF4.

* Laufende Nr.: Familie im Haushalt (trad. Familienkonzept): EF25

* Einschränkung der Variablen EF31 "Bevölkerung: Privathaushalte".

* 1 "Bevölkerung in Privathaushalten".

* 2 "Bevölkerung in Gemeinschaftsunterkünften".

* M0 "Entfällt /Haushalt ist leer oder nicht.

* befragt)".

* auf die Bevölkerung in Privathaushalten (EF31=1).

* EF34 "Stellung innerh. der Familie (tradit.
* Familienkonzept)".

* 1 "Bezugsperson in der Familie".

* 2 "Ehefrau der Bezugsperson in der Familie".

* 3 "Lediges Kind in der Familie".

* M 0 "Entfällt (Bevölkerung in.

* Gemeinschaftsunterkünften).

* Einschränkung der Variablen EF35 "Stellung innerhalb der Familie/Lebensform".

* 1 "Bezugsperson in der Familie/Lebensform".

* 2 "Ehefrau oder Lebenspartner(in) der. Bezugsperson in der Familie/Lebensform".

* 3 "Lediges Kind in der Familie".

* M0 "Entfällt (Bevölkerung in.

* Gemeinschaftsunterkünften)".*/

sort cases by EF1 EF3 EF4 EF25.

select if EF31=1.

compute x866=0.

if (EF31=1 & EF34=3 or EF35=3) x866=1.

aggregate outfile=*

mode=addvariables

/presorted

/break EF1 EF3 EF4 EF25

/v866'Replikation: Zahl der ledigen Kinder insgesamt in der Familie'= sum(x866).

crosstabs ef866 by v866

/missing include.